Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

DTIC ELECTE MAR 29 1990 D

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS: (not needed)

2a. SECURITY CLASSIFICATION AUTH: None

3. DISTRIBUTION/AVAILABILITY OF REPORT: Unclassified/ unlimited

(2)

AD-A219 930

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 90-0341

6a. NAME OF PERFORMING ORGANIZATION: University of Missouri-Columbia

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State, and ZIP Code): Columbia, MO 65211

7b. ADDRESS (City, State, and ZIP Code): Building 410, Bolling AFB, Washington D.C. 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR/NM

8b. OFFICE SYMBOL (If applicable):

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-87-0226

8c. ADDRESS (City, State, and ZIP Code): Building 410, Bolling AFB, D.C. 20332-6448

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2304	A7	

11. TITLE (Include Security Classification): Management of Uncertainty in Military Scene Analysis

12. PERSONAL AUTHOR(S): Keller, James, M. Crownover, Richard, M. McLaren, Robert, W.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 87/6/1 TO 88/5/31	88/7/26	193

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Uncertainty, Fuzzy Logic, Belief Theory, Fuzzy Integral, Rule Based Automatic Target Recognition, Fractal Geometry, Texture Analysis, Linear Descriminant Analysis, Context

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Research into modeling and managing uncertainty in military scene analysis is presented. A new method of logical inference was developed for the case where propositions are modeled by possibility distributions. This scheme was tested in a prototype fuzzy rule-based system for automatic target recognition. An information fusion technique based on the fuzzy integral was also developed. This was inserted into a numeric uncertainty propogation ATR prototype system.

Fractal geometry was exploited for scene description and segmentation. Results concerning dimension calculation, texture description and segmentation, and surface orientation from fractal features is presented.

Fast solutions to two problems in linear discriminant analysis are contained herein. Both techniques avoid the potentially disastrous errors from calculating large cross-product matrices.

Finally, preliminary work on modifying confidence values for hypotheses based on external or scene derived context are presented.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
Dr. Abraham Waksman	(202) 767-5027	NM

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

MANAGEMENT OF UNCERTAINTY IN MILITARY SCENE ANALYSIS

James M. Keller*, Richard M. Crownover**, and Robert W. McLaren*

* Electrical and Computer Engineering Department
** Mathematics Department
University of Missouri-Columbia
Columbia, MO 65211

Final Report for grant AFOSR-87-0226 for the period
1 June 1987-31 May 1988

Prepared for:

Air Force Office of Scientific Research
Building 410
Bolling Air Force Base
Washington D.C. 20332-6448

TABLE OF CONTENTS

1.0 Introduction

This report documents the research in the management of uncertainty in military scene analysis performed at the University of Missouri-Columbia under the Air Force Office of Scientific Research grant number AFOSR-87-0226 during the period June 1, 1987 to May 31, 1988. The research accomplishments are summarized in section 2. These topics are then expanded upon in subsequent sections. Our research on uncertainty modeling is developed in section 3. Section 4 describes the results obtained using fractal geometry for region description. In section 5, we present a new method for a fast least squares approach to linear discriminant analysis. Two rule-based systems for Automatic Target Recognition (ATR), one incorporating uncertainty using the fuzzy integral and Dempster-Shafer belief theory and the other using fuzzy logic, are described in section 6. Finally, in section 7, we present the preliminary results on the use of external context in modifying the confidence results from the numeric-based ATR system.

The breadth of the research was made possible by the interdisciplinary nature of the UMC research team and the close cooperation of Emerson Electric Electronics and Space Division in St. Louis. We have been performing ATR-related computer vision research in Emerson Electric for five years and this relationship has strengthened the quality of the research described herein.

2.0 Summary of Accomplishments

The goal of this research, as stated in the proposal, was to perform the basic research necessary to model and effectively handle uncertainty present in military scenes for the detection and recognition

of image regions and objects. Six papers, two Ph.D. dissertations and one M.S. Thesis resulted from the research performed under this grant. Two more papers stemming from the dissertations are currently under preparation.

In order to accomplish the above stated goal, several thrusts were pursued. We approached the modeling of uncertainty in computer vision from two directions. In the first approach, we modeled the uncertainty in a proposition numerically. We developed a new nonlinear information fusion technique, the fuzzy integral, to combine low level objective information from object features with the expectation of importance of these features towards object classification [1,2]. These values were used to create support functions in the Shafer sense [3], and the results from different algorithms and rules were combined using Dempster's Rule in an ATR system [4]. The alternate direction involved the development of a new fuzzy logic inference mechanism [5]. This new technique which uses the concept of truth value restriction was shown to be superior to ten accepted fuzzy logic inference schemes under a variety of conditions [5]. We built a fuzzy logic rule-based system for ATR and tested it on the same data used for the numeric scheme with excellent results. The results of both systems will be described in subsequent sections.

During this past year, we performed research into the description of natural scene regions using fractal geometry [6-8]. This involved extensions of our previous work to surfaces [9]. The invariance of fractal dimension coupled with the sensitivity of our new parameter, the average Holder constant, have resulted in new algorithms for determination of distance and orientation of fractal regions [8].

Our new implementation of lacunanty [6,8] has produced excellent texture description and segmentation.

In any pattern recognition problem, the dimensionality of feature space poses problems for computation. Linear discriminant analysis is one technique to reduce the dimensionality while preserving the separability of the data. We have used these methods successfully in an earlier contract from Emerson Electric to perform target detection and recognition using vectors of gray levels from image windows [10,11]. In [12] we developed methods for fast solutions to two problems in a least squares sense. Both techniques avoid potentially disastrous errors from calculating large cross-product matrices.

The methods developed for modeling and manipulating uncertainty in military scene analysis were incorporated into rule-based systems [4,5]. These systems were tested with data extracted from sequences of FLIR images. The results demonstrated the flexibility and inherent advantages of such approaches. Finally we incorporated a preliminary methodologies for modifying the confidence values generated by both systems based on external context [13]. These approaches show considerable promise, and more work is currently underway.

Each section in this report is self contained in that references, figures, and tables are all included within the section for ease of reading among the different topics.

REFERENCES

1. H. Tahani, "Information Fusion in Computer Vision Using the Fuzzy Integral", M.S. Thesis, University of Missouri-Columbia, 1987.

2. H. Tahani and J. Keller, "The Fusion of Information in Computer Vision Using the Fuzzy Integral", IEEE Trans. Syst. Man and Cybern., under review.

3. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

4. J. Wootton, J. Keller, C. Carpenter, and G. Hobson, "A Multiple Hypothesis Rule-Based Automatic Target Recognizer", in Pattern Recognition, Lecture Notes in Computer Science, Vol. 301, J. Kittler (ed.), Springer-Verlag, 1988, pp 315-324.

5. A. Nafarieh, "A New Approach to Inference in Approximate Reasoning and its Application to Computer Vision", Ph.D. Dissertation, University of Missouri-Columbia, 1988.

6. J. Keller, S. Chen, and R. M. Crownover, "Texture Description and Segmentation Through Fractal Geometry", Computer Vision, Graphics, and Image Processing, accepted for publication.

7. S. Chen, J. Keller, and R. Crownover, "On the Calculation of Fractal Features from Images", Second Int. Conf. on Computer Vision, Tarpon Springs, FL, Dec. 1988, under review.

8. S. Chen, "Fractal Geometry in Image Understanding", Ph.D. Dissertation, University of Missouri-Columbia, 1987.

9. J. Keller, R. Crownover, R. Chen, "Characteristics of Natural Scenes Related to the Fractal Dimension", IEEE Trans. Pattern Anal. and Machine Intell., Vol. 9, No. 5, 1987, pp 621-627.

10. J. Keller, R. Crownover, J. Wootton, and G. Hobson "Target Recognition Using the Karhunen-Loeve Transform", Proc. Int. Conf. on Syst. Man and Cybern., Tucson AZ, Nov. 1985, pp 310-314.

11. G. Hobson, J. Keller and R. Crownover, "Method of Extracting Uncorrelated Features from Gray Level Images for Target Recognition", Proc. 8th Annual Symposium on Ground Vehicle Signatures, Vol. I, Houghton, MI, Aug 1986, pp 208-214.

12. R. Crownover, "A Least Squares Approach to Linear Discriminant Analysis", SIAM J. Scientific. Stat Comput., under review,

13. A. Mogre, R. McLaren, and J. Keller, "Confidence Modification Using Context in Computer Vision", Proc. SPIE Symposium on Intelligent Robots and Computer Vision, Cambridge, MA, Nov. 1988, to appear.

3. Modeling of Uncertainty in Computer Vision

In our research on modeling uncertainty in military scene analysis, we took both numeric and set-based approaches. On the numeric side, we implemented a Dempster-Shafer belief structure in a rule-based ATR (explained in section 6) and developed an information fusion technique around the fuzzy integral. On the other hand, we developed a new fuzzy logic inference scheme and built a prototype ATR scene analysis system with 50 rules. In this section we describe the research based of fuzzy set theory. The rule-based systems are presented in section 6.

3.1 Fuzzy Logic in Computer Vision

Fuzzy sets were introduced by Zadeh in 1965 [1]. Since that time, researchers have found numerous ways to utilize this theory to generalize existing techniques and to develop new algorithms in pattern recognition, decision analysis and risk analysis [2-10]. Fuzzy sets generalized the traditional membership of an element in a set from the binary (0,1) to a value in the interval [0,1]. Most traditional or crisp set theoretic operations have analogs in fuzzy set theory [11]. We have developed both pattern recognition algorithms and segmentation techniques which incorporate membership information into the final decision [6-10,12].

Possibility distributions [4-5] form the basis for fuzzy logic. If Y is a variable which takes values in a universe of discourse U, then a possibility distribution associated with Y may be viewed as an elastic constraint on the values that may be assigned to Y. For example, if F is a fuzzy subset of U characterized by its membership function

μ_F:U--→[0,1], then the statement "Y is F" translates into a possibility distribution for Y being equal to F. In particular, we may write

Possibility (Y = u) = μ_F(u).

In this way we are able to effectively model conditions in real military scenes such as:

roads are usually straight and thin,

treelines are rugged,

background clutter is high,

features a, b, c work well at night to describe tanks.

Fuzzy logic has been developed to provide decision making capabilities in the presence of uncertainty [4-5]. Its structure is rule based. However, in this case, the uncertainty in statements and conditions is modeled as possibility distributions. The antecedent clause, the consequent clause or both, may be represented as possibility distributions. As an example, we may have a rule such as

IF the region is straight and thin,

THEN the region is a ROAD;

or more generally

IF the region is straight and thin,

THEN confidence in the class ROAD is high.

In this example, straight, thin, and high are modeled by possibility distributions over appropriate domains: straight may be defined as a fuzzy set in terms of average curvature, thin by the diameter of the region, and high by a fuzzy set over a closed interval of reals. A system of inference, called approximate reasoning, has been developed to

make deductions from statements expressed in terms of possibility distributions.

The original fuzzy inference mechanism extended the traditional modus ponens rule which states that from the propositions

P_1: If S is A Then Y is B

and P_2: X is A,

we can deduce Y is B. If proposition P_2 did not exactly match the antecedent of P_1, for example, X is A', then the modus ponens rule would not apply. However, in [5], Zadeh extended this rule if A, B, and A' are modeled by fuzzy sets, as suggested above. In this case, P_1 is characterized by a possibility distribution.

$$\Pi_{(X|Y)} = R \text{ where}$$

$$\mu_R(u,v) = \min \{1, \max \{(1-\mu_A(u)), \mu_B(v)\}\}.$$

It should be noted that this formula corresponds to the statement "not A or B", the logical translation of P_1. Zadeh now makes the inference Y is B' from μ_R and $\mu_{A'}$ by

$$\mu_{B'}(v) = \max_u \{\min \{\mu_R(u,v), \mu_{A'}(u)\}\}.$$

While this formulation of fuzzy inference directly extends modus ponens, it suffers from several problems [13,14]. In fact, if proposition P' is "X is A," the resultant fuzzy set is not exactly the fuzzy set B. Several authors [13-16] have performed theoretical investigations into alternative formulations of fuzzy implications.

Besides changing the way in which P_1 is translated into a

possibility distribution, methods involving truth modification have been proposed. In this approach, the proposition X is A' is compared with X is A, and the degree of compatibility is used to modify the membership function of B to get that for B'.

During this past year, we have developed a new scheme for truth value restriction based on a novel compatibility measure between the fuzzy set A and A' [13]. In this methodology, we define

$$\text{comp}\ (A,\ A') = \frac{|A \cap A'|}{|A \cup A'|}$$

where $|\cdot|$ denotes the area under the fuzzy set. This formulation of the compatibility retains all the information in A and A'.

From this compatibility, a truth restriction was obtained [13] and the result of the fuzzy implication was generated. We have proved that this technique provides the intuitively correct exact results under reasonable hypotheses, and that it outperforms the other approaches in numerous simulation studies. Table 1 gives the fuzzy set definitions for a simple set of linguistic terms used in one simulation study. Table 2 shows intuitive relations which should exist when the inputs may not exactly match the rule but are given as functions of the antecedent. Table 3 gives the percent error obtained for several values of antecedent for the ten standard operators and our new scheme (labeled "proposed"). Note that in all cases the new inference mechanism produced the correct result with no error. Also shown is the result of Modus Tollens on the same rule, where again, the new method outperforms the other ten.

Complete results of the various simulation studies can be found in [13], the body of which is included in the appendix. In section 6 we

apply this new reasoning approach to a prototype fuzzy logic production system for an automatic target recognition problem.

Table 1. The meaning of linguistic terms over the domain [1,11] sampled at integer points.

Name	Membership										
small	1.00	0.67	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
very small	1.00	0.45	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
morl small	1.00	0.82	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
not small	0.00	0.33	0.57	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
medium	0.00	0.00	0.25	0.50	0.75	1.00	0.75	0.50	0.25	0.00	0.00
very medium	0.00	0.00	0.06	0.25	0.56	1.00	0.56	0.25	0.06	0.00	0.00
morl medium	0.00	0.00	0.50	0.71	0.87	1.00	0.87	0.71	0.50	0.00	0.00
not medium	1.00	1.00	0.75	0.50	0.25	0.00	0.25	0.50	0.75	1.00	1.00
high	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.40	0.60	0.80	1.00
very high	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.16	0.36	0.64	1.00
morl high	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.63	0.77	0.89	1.00
not high	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.60	0.40	0.20	0.00
unknown	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

morl = more or less

Table 2. Some intuitive relations between X and Y in proposition "If X is B then Y is C".

Relation	If	Then
I (modus ponens)	X is B	Y is C
II	X is very B	Y is very C
III	X is more or less B	Y is more or less C
IV	X is not B	Y is unknown
V (modus tollens)	Y is not C	X is not B

Table 3. Percentage error in applying different operators on implication "If X is small then Y is High" for several values of X and Y.

Operator	% error when X is				% error when Y is
	small	very small	morl small	not small	not high
1	41	57	53	0	5
2	41	57	53	0	5
3	41	57	53	11	5
4	0	27	20	86	93
5	47	58	51	0	8
6	0	20	15	0	6
7	22	24	21	0	2
8	63	65	60	0	10
9	46	58	51	0	8
10	47	58	51	0	8
proposed	0	0	0	0	0
expected result	high	very high	morl high	unknown	not small

morl = more or less

3.2 The Fuzzy Integral and Information Fusion

In this section we describe the research involving the fuzzy integral and confidence generation in military application of computer vision. Full details can be found in [17], which is included in the appendix.

Information fusion is an important aspect of any intelligent system. The rationale behind bringing multiple input information sources is that the information that we have to deal with in each source is either partial or contaminated, that is, it is uncertain and/or imprecise [18-20]. An example of such a system is the following: In the field of computer vision the goal of image understanding is the design and implementation of a system which will be able to determine the mapping from what is actually sensed, the image, to the scene of interest. The system must be able to determine what objects are in the scene and in what spatial relationship they lie. Unfortunately, the problem is vastly underconstrained, in general, since images are ambiguous and can be the result of an infinite number of scenes. Now, by adding multiple input information sources ambiguities may be resolved because similar aspects of the scene will be encoded in different ways by different information sources.

The question of certainty in a representation or decision is, however, a question of evidence. Each source of information can be considered as a body of evidence for supporting or rejecting a decision or hypothesis. The task is to combine this evidence to make a final decision. We now define the fuzzy integral as an evidence combination scheme.

Definition 1. Let X be a non-empty set, β be a σ-algebra of X. A fuzzy measure is a real valued set function g defined on β satisfying the following properties:

(1) $g(\Phi) = 0$, $g(X) = 1$;
(2) If $A, B \in \beta$ and $A \subset B$, then $g(A) \leq g(B)$;
(3) If $\{A_n\} \in \beta$ such that $A_1 \subset A_2 \subset \ldots$,

then $g(\bigcup_{i=1}^{\infty} A_i) = \lim_{i \to \infty} g(A_i)$.

We note that fuzzy measure generalizes probability measure in that it does not require additivity. A particularly useful set of fuzzy measures is due to Sugeno [21].

Definition 2. Let g_λ be a fuzzy measure satisfying the addition property:

If $A \cap B = \Phi$, then $g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda\, g_\lambda(A) g_\lambda(B)$, for some $\lambda > -1$.

Then g_λ is called a Sugeno measure.

Suppose X is a finite set, $X = \{x_1, \ldots, x_n\}$, and let $g^i = g_\lambda(\{x_i\})$. Then the set $\{g^1, \ldots, g^n\}$ is called the fuzzy density function for g_λ. Using the above definitions one can easily show that g_λ can be constructed from a fuzzy density function by

$$g_\lambda(A) = [\prod_{x_i \in A}(1 + \lambda g^i) - 1]/\lambda,$$

for any subset A of X. Using the fact that $X = \bigcup_{i=1}^{n} \{x_i\}$, λ can be

determined from the equation

$$1 = [\prod_{i=1}^{n} (1 + \lambda g^i) - 1]/\lambda \ . \qquad (1)$$

Definition 3. Let H: X→ [0,1]. The fuzzy integral of h over X with respect to g_λ is defined in [21] by:

$$\int_X h(x) \circ g_\lambda = \sup_{\alpha \in [0,1]} [\alpha \wedge g_\lambda (F_\alpha)]$$

where $F_\alpha = \{x \in X \mid h(x) \geq \alpha\}$.

If X is a finite set, $X = \{x_1, \ldots, x_n\}$, arranged so that $h(x_1) \geq h(x_2) \geq \ldots \geq h(x_n)$, then

$$\int_X h(x) \circ g_\lambda = \bigvee_{i=1}^{n} [h(x_i) \wedge g_\lambda (X_i)] \qquad (2)$$

where $X_i = \{x_1, \ldots, x_i\}$. Also, given λ as calculated above, the values $g_\lambda(X_i)$ can be determined recursively as

$$g_\lambda(X_1) = g_\lambda(\{x_1\}) = g^1; \qquad (3a)$$

$$g_\lambda(X_i) = g^i + g_\lambda(X_{i-1}) + g^i g_\lambda(X_{i-1}) \qquad (3b)$$

for $1 < i \leq n$.

The fuzzy integral is interpreted as a subjective evaluation of objects where the subjectivity is embedded in the fuzzy measure. In comparison with probability theory, the fuzzy integral corresponds to the concept of expectation [21]. In general, fuzzy integrals are nonlinear

functionals (although monotone) whereas ordinary (eg, Lebesque) integrals are linear functionals. It is this nonlinear subjective evaluation potential of the fuzzy integral which we utilize in the fusion of different information sources.

The calculation of the fuzzy integral with respect to a g_λ fuzzy measure only requires the knowledge of the density function, where the $i^{\underline{th}}$ density, g^i, may be interpreted as the degree of importance of x_i for $i = 1,2,\ldots,n$. The degree of importance, furthermore may be interpreted as a belief function if

$$\sum_{i=1}^{n} g^i < 1,$$

and a plausibility if this sum is greater than 1.

The fuzzy integral was used as a segmentation tool in [9,10]. Here, the design and the implementation of an object recognition system using the fuzzy integral will be explained. The output of this system can be considered as a decision, or a hypothesis for a higher level of recognition.

In many cases, an object can be represented as a vector in an n-dimensional Euclidean space, where each component of this vector is a feature measured from that object. There are many different types of features that can be calculated from objects, e.g. shape measures, texture measures, and statistical measures, to name a few. The reason for measuring different features is that there is usually no single feature that can identify the objects of interest. In fact, there is

normally no set of features which always distinguishes an object from others precisely. There is always an uncertainty inherent in the recognition problem. Instead, each feature or group of features can be considered as evidence in the identification of an object. Obviously, each of these features or group of features would have a degree of importance in the identification of an object.

Let X be an object described by n features, $X = (x_1, \ldots, x_n)$. For each pattern class w_j, let $u_j: X \to [0,1]$. Thus, u_j is an objective partial evaluation of X from class w_j, that is, for each feature x_i, $\mu_j(x_j)$ measures the membership of X in w_j from the standpoint of a single feature x_i. This partial evaluation is combined with the subjective measure $g_{\lambda j}$ which represents the important degree of the subset $X_i = (x_1, \ldots, x_i)$ of X. For example, $g_{\lambda j}(X_1) = g_{\lambda j}((x_1))$ expresses the extent to which a viewpoint of feature x_1 is important in evaluating objects from class w_j, and for $i > 1$, $g_{\lambda j}(X_i) = g_{\lambda j}((x_1, \ldots, x_i))$ expresses the degree to which the set of viewpoints $(x_1, \ldots, x_i)$ contribute to the recognition of objects from class w_j. The fuzzy integral value,

$$e_j = \bigvee_{i=1}^{n} [u_j(x_i) \wedge g_{\lambda j}(X_i)], \qquad (4)$$

gives a nonlinear evaluation of the degree to which object X belongs to class w_j.

The Sugeno measure $g_{\lambda j}$ for each class is generated from a fuzzy density function $(g_j^{\,i}, \ldots, g_j^{\,n})$ by equation 3. The densities for each feature can be obtained subjectively from an expert or can be generated from a set of training data. The attractiveness of this is that one

need only consider a single feature (or small group of features) to fix the density function from which the entire fuzzy measure is calculated.

This same approach is used to combine information from different algorithms, different sensors or information over time. The fuzzy integral is a very general paradigm in this regard.

Results of this algorithm on simulation data can be found in [17]. Here we highlight the application to military imagery. The data consisted of several sequences of FLIR images containing an armored personnel carrier (APC) and two different tanks. A pre-processing step was run on each image to detect objects of interest, and features were calculated for each of these objects.

The feature level integration was performed using four statistical features. To get the partial evaluation, h(x), for each feature we used the fuzzy 2-mean algorithm [2]. The fuzzy densities, the degree of importance of each feature, were assigned subjectively based on how well these features separated the two classes Tank and APC on training data. The result is presented in the form of confusion matrix, in Table 4, where the count of samples listed in a row are those which belong to the corresponding class and the count of samples listed in each column are those after classification. As can be seen, the fuzzy integral performed better than Bayes of this data set.

Next three classifiers, the fuzzy integral, the Bayes classifier and the fuzzy perceptron were applied to the data. The a posteriori probabilities obtained from the Bayes classifier together with classification results from fuzzy integral and the fuzzy perceptron were taken as partial evaluations for the objects of interest. The degree of importance was subjectively based on how good these classifier performed

on a training data set. Then these bodies of evidence were combined using the fuzzy integral. The results are in Table 5. The results show that this methodology produces good estimates of class confidence based on the objective information and the subjective expectation of the importance of this information. As can be seen, objects 13 and 14 (APC's) were misclassified by the Bayes algorithm. However, in the final evaluation, they were correctly classified. The effect of misclassification by Bayes has given rise to small fuzzy integral values for the APC hypothesis in both cases. This information can be used by an intelligent monitor to initiate more sophisticated procedures to increase confidence in class membership.

Table 4a

The result of the feature level fuzzy integral

	g^1	g^2	g^3	g^4	λ
Tank	0.1	0.21	0.2	0.3	0.736
APC	0.08	0.15	0.17	0.24	2.022

		Tank	**APC**
Tank	100.0%	176	0
APC	71.2%	19	47

Total Correct: 92.15%

Table 4b

The result of the Bayes classifier

		Tank	**APC**
Tank	100.0%	176	0
APC	66.7%	22	44

Total Correct: 90.91%

Table 5a

The result of information fusion using the fuzzy integral on three different classifiers for Tank

Actual Object Class	Partial evaluation for Tank			Fuzzy Integral Evaluation for Tank Hypothesis
	Bayes	Fuzzy K-mean	Feature level Fuzzy integral	
1 Tank	1.00	0.77	0.68	0.68
2 Tank	1.00	0.85	0.71	0.71
3 Tank	1.00	0.81	0.71	0.71
4 Tank	1.00	0.83	0.71	0.71
5 Tank	1.00	0.76	0.71	0.71
6 Tank	1.00	0.78	0.66	0.66
7 Tank	1.00	0.83	0.66	0.66
8 Tank	1.00	0.78	0.68	0.68
9 Tank	1.00	0.73	0.64	0.64
10 APC	0.44	0.44	0.40	0.40
11 APC	0.00	0.27	0.27	0.27
12 APC	0.00	0.53	0.49	0.43
13 APC	0.99	0.26	0.25	0.26
14 APC	0.97	0.18	0.21	0.21
15 APC	0.00	0.23	0.27	0.23
16 APC	0.00	0.23	0.24	0.23
17 APC	0.00	0.29	0.28	0.28
18 APC	0.00	0.29	0.26	0.26

Table 5b

The result of information fusion using the fuzzy integral on three different classifiers for APC

Actual Object Class	Partial evaluation for Tank			Fuzzy Integral Evaluation for APC Hypothesis
	Bayes	Fuzzy K-mean	Feature level Fuzzy integral	
1 Tank	0.00	0.23	0.32	0.23
2 Tank	0.00	0.15	0.28	0.20
3 Tank	0.00	0.19	0.29	0.20
4 Tank	0.00	0.17	0.28	0.20
5 Tank	0.00	0.24	0.27	0.24
6 Tank	0.00	0.22	0.34	0.22
7 Tank	0.00	0.17	0.34	0.20
8 Tank	0.00	0.22	0.32	0.22
9 Tank	0.00	0.27	0.36	0.27
10 APC	0.56	0.56	0.55	0.55
11 APC	1.00	0.27	0.73	0.72
12 APC	1.00	0.49	0.47	0.47
13 APC	0.01	0.75	0.65	0.33
14 APC	0.03	0.82	0.65	0.33
15 APC	1.00	0.77	0.59	0.24
16 APC	1.00	0.77	0.65	0.65
17 APC	1.00	0.72	0.65	0.65
18 APC	1.00	0.74	0.65	0.65

REFERENCES

1. L. A. Zadeh, "Fuzzy Sets," Inf. Control, Vol. 8, pp 338-353, 1965.

2. J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

3. M. Gupta, R. Ragade and P. Yager, eds., Advances in Fuzzy Set Theory and Applications, North-Holland, Amsterdam, 1979.

4. L. A. Zadeh, "The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems," Fuzzy Sets and Systems, Vol. 11, No. 3, 1983, pp. 199-228.

5. L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning I, II, III," Information Sciences, Vol. 8, No. 3, 1975, pp. 199-249, Vol. 8, No. 4, 1975, pp. 301-357, Vol. 9, No. 1, 1976, pp. 43-80.

6. J. Keller, and D. Hunt, "Incorporating Fuzzy Membership Functions into the Perceptron Algorithm," IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. PAMI-7, No. 6, November 1985, pp. 693-699.

7. J. Keller, M. Gray, and J. Givens, "A Fuzzy K Nearest Neighbor Algorithm," IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, July/August, 1985, pp. 580-585.

8. J. Keller, and J. Givens, "Membership Function Issues in Fuzzy Pattern Recognition," Proceedings, International Conference on Systems, Man, Cybernetics, Tucson, AZ, November 1985, pp. 210-214.

9. J. Keller, H. Qiu, and H. Tahani,"The Fuzzy Integral in Image Segmentation," Proceedings, NAFIPS-86, New Orleans, June 1986, pp. 324-338.

10. H. Qiu, and J. Keller, "Multispectral Segmentation Using Fuzzy Techniques, " Proceedings NAFIPS-87, Purdue University, May 1987, pp. 374-387.

11. D. Dubois and H. Prade, Fuzzy Sets and Systems: Theory and Applications, Academic Press, New York, 1979.

12. J. Keller, and C. Carpenter, "Image Segmentation in the Presence of Uncertainty," Proceedings NAFIPS-88, San Francisco, CA, June 1988, pp 136-140.

13. A. Nafarieh,"A New Approach to Inference in Approximate Reasoning and its Application to Computer Vision," Ph.D. Dissertation, University of Missouri-Columbia, 1988.

14. M. Mizumoto, S. Fukami, and K. Tanaka, "Some methods of fuzzy reasoning," in Advances in Fuzzy Set Theory and Applications, Gupta, M. M., Ragade, R. K., and Yager, R. R., (Eds.), Ansterdam, The Netherlands: North-Holland, 117-136, 1979.

15. M. Mizumoto, "Fuzzy reasoning with 'if ... then... else ...'," in Applied Systems and Cybernetics, G. E. Lasker, (Ed.), New York: Pergamon, 2927-2932, 1981.

16. J. F. Baldwin, and N. C. Guild, "Feasible algorithms for approximate reasoning using fuzzy logic," Fuzzy Sets and Systems, Vol. 3, 225-251, 1980.

17. H. Tahani and J. Keller, "The Fuzzy Integral and Information Fusion," IEEE Trans. Syst. Man. Cybern, under review.

18. E. M. Riseman and A. R. Hanson, "A methodology for the Development of General Knowledge-Based Vision Systems," Proceeding of the IEEE Workshop on Principle of Knowledge-based Systems, pp. 159-170, Dec. 1984.

19. A. R. Hanson and E. M. Riseman, "VISION: A Computer System for Interproeting Sciences," In Computer Vision Systems, eds. A. R. Hanson and E. M. Riseman, Academic Press, NY, 1978.

20. B. E. Flinchbaugh and B. Chandrasekaran, "A Theory of Spatio-Temporal Aggregation for Vision," Artificial Intelligence. Vol. 17, pp. 387-407.

21. M. Sugeno, "Fuzzy Measures and Fuzzy Integrals: A survey," In Fuzzy Automata and Decision Processes, North-Holland Publ., Amsterdam, pp. 89-102, 1977.

4.0 Fractal Geometric Scene Characteristics

One of the tasks this past year was to study fractal geometry and its applications to computer vision. Our goal is to develop robust parameters for region description and segmentation. We made significant progress in texture description and segmentation, improved calculation of fractal dimension, and surface orientation from fractal parameters.

Fractal geometry provides useful models for describing complex surfaces and curves in images of natural scenes. It is important to be able to identify background clutter for optimal performance of target recognition algorithms and it is important to be able to ascertain distance scales and shape of terrain in these images. Our recent work [1-4] on the use of fractal models introduces new concepts and approaches for recognizing fractal objects, distances of these objects from the image plane, texture segmentation from fractal features, and recovering shape from fractality.

A fractal is a geometric configuration having Hausdorff dimension (usually a fraction) that is greater than its topological dimension. The fractals of natural scenery are statistically self-similar, in that any part can be decomposed into a certain number N of copies that are scaled by a factor of r having the same statistical properties. The parameters N and r are related to the fractal or self-similarity dimension by the equation

$$Nr^d = 1$$

or

$$d = \log(N)/\log(1/r).$$

B. Madelbrot, who introduced the term fractal, has written a diverse casebook on fractals in nature [5]. In addition, Mandelbrot and Van Ness [6], introduced the fractional Brownian motions, in terms of which most of the fractal models we have utilized thus far are expressed. A fractional Brownian motion (fBm) is a generalization of classical Brownian motion (in one or more variables). An important property of a fBm B(t) is that the increments are Gaussian normal, satisfying

$$Pr(\{B(t + T) - B(t)/ T^{H} < x\} = erf(x),$$

where H is a parameter relating the fractal dimension d of the graph of B by d = 2 - H for the one-variable case or d = 3 - H for the two-variable case. The standard deviations of the increments satisfy a power law

$$\sigma(B(t + T) - B(t)) = c\ T^{H},$$

which is important in determining the parameter H, and hence the fractal dimension d. This is done by plotting log (σ) vs. log (T) and performing linear regression analysis.

In a series of papers, including [7,8], Pentland has presented evidence that most natural surfaces are spatially isotropic fractals and that their intensity images are also fractals. In arguing the suitability of the fractal model, Pentland set forth a methodology to compute the fractal dimension using the fBm model and to use this evaluation to perform texture segmentation and classification.

Medioni and Yasumoto [9] conducted further research on segmenting natural scenes using fractal parameters. They concluded that fractal dimension alone cannot separate textures that differ in roughness. Peleg et al. [10], using still other methods, reported good classification results and that fractal measurements can prove helpful in characterizing texture. A maximum likelihood estimator was developed by Lundahl et al. [11] to estimate the fractal dimension related to parameter H. Their work was applied to X-ray images and indicated strong potential for quantifying texture.

A parameter related to H, the average Holder constant, is defined as

$$\bar{\alpha} = \text{avg}\{\log|B(t+T) - B(t)|/\log(T)\}.$$

For small increments, $\bar{\alpha} = H$. In [1-3], we derived a more useful higher order relation. We showed that for a fBm, the average Holder constant in the one-variable case satisfies

$$\bar{\alpha} = H + c/\log(T). \qquad (1)$$

If the graph of a fBm is scaled by a factor of s, or if an intensity image of such a graph is scaled, then the average Holder constant $\bar{\alpha}_s$ for the scaled version satisfies

$$\bar{\alpha}_s = c_1 \log(s) + c_2. \qquad (2)$$

Thus the average Holder constant is scale sensitive in a recoverable way. Equivalently, the average Holder constant is sensitive to the distance of an object from the image plane. This property allowed us to make distance estimates of tree silhouettes and mountain

silhouettes in images. Figure 1 shows a plot of the average Holder constant verses the log of the scale factor for a sequence of images of a tree scene at different scales. The regression line is also included to demonstrate the linearity of the average Holder constant with respect to $\log(\lambda)$. In [1] this was used to predict the scale of an image. In addition, an initial study at UMC has been completed by Chen [4] on recovering the orientation of a plane that is a fractal surface. Chen used perspective geometry of images as described by Ohta et al. [12] in addition to properties of average Holder constants.

The above estimates require that the fractal be modeled by a fractional Brownian motion. More generally, the dimension of a fractal set can be calculated by the box dimension. The box dimension of a self-similar fractal is defined in terms of the number of boxes N(L) of side L that it takes to cover the set. The quantity N(L) is related to the fractal dimension by

$$N(L) = c/L^d$$

and the value of d is determined from linear regression analysis from the plot of log(N(L)) vs. log(L).

Voss [13] introduced a sharper calculation of the box dimension in which he used the parameter P(m,L) = Pr(m points lie in a box of side L centered at an arbitrary point in the fractal set). If M is the number of points in the image and N is the number of points in a box of side L, then

$$N(L) = \sum_{m=1}^{N} \left(\frac{M}{m} \right) P(m,L),$$

and thus

$$\sum_{m=1}^{N} \frac{1}{m} \; P(m,L) = c/L^{d}.$$

Again, linear regression for a log-log plot is used to determine the dimension d. If box sizes are too small, these box dimension calculations are in error due to the sparsity of boxes relative to the curve to be covered. In our papers [2,4], the extent of this error is investigated and lower bounds for box sizes that give correct results are explored through simulations. A new method of calculating the box dimension is presented, based on interpolating the curve or surface linearly and adding more boxes so as to completely cover the interpolated curve. It is shown that this new method significantly corrects the deficiencies of the previous methods. As an example, figure 2 shows the histogram of estimated fractal dimensions (as described in [9]) of a fractal mosaic image comprised of regions with fractal dimensions 2.2, 2.4, and 2.6, each region generated by the Fourier Spectrum method [7]. The dimension was estimated for windows of 16 x 16 pixels with a movement of 4 pixels between windows. It is difficult, if not impossible, to determine two good thresholds to segment this image. The straight implementation of box dimension produced worse results. Figure 3 shows the corresponding histogram. However, the new interpolation method produced the histogram shown in figure 4. As can be seen this approach yielded excellent separability in the composite image. More details of this method can be found in [2], and which is included in the appendix. Figure 5a shows the fractal composite image, whereas Figure 5c and 5d shows the segmentation of this image by the interpolated and non-interpolated box dimension. The

segmentation was performed by the K-Means clustering algorithm. The interpolated box dimension alone produced excellent results.

Mandelbrot [5] and Voss [13] have introduced lacunarity (gap) measurements to distinguish fractals that have the same dimension. Mandelbrot's lacunarity measurement is

$$\Lambda = E\left[\left(\frac{W}{E(W)} - 1\right)^2\right],$$

where W is mass and E(•) is expected value. Voss's lacunarity is defined in terms of

$$M(L) = \sum_{m=1}^{N} mP(m,L)$$

and

$$M^2(L) = \sum_{m=1}^{N} m^2 P(m,L),$$

and is given by

$$\Lambda(L) = \frac{M^2(L) - [M(L)]^2}{(M(L))^2}$$

In [3,4] we introduced a new lacunarity measurement,

$$C(L) = \frac{M(L) - N(L)}{M(L) + N(L)},$$

where

$$N(L) = \sum_{m=1}^{N} \frac{1}{m} P(m,L),$$

which like the others, is a second order statistical property of the mass distribution. It was found that the new lacunarity measurement gives improved texture segmentation compared to the previous methods and that the best texture segmentation results from using feature vectors

consisting of the fractal dimension d and the lacunarity measurements C(L) for several, say 4, values of L. Such segmentations significantly improve performance over those using just the fractal dimension. Figure 5(b) displays the results of segmentation of the fractal composite using interpolated fractal dimension and 4 lacunarity values. Note that the only classification errors occur at the region boundaries. However, as should be expected for artificially generated fractal surfaces, there is little improvement over using just our new estimate of dimension.

For texture images, however, the situation is different. Figure 6(a) displays 4 natural textures: pigskin, grass, sand, and raffia. Figure 7 gives the histograms of fractal dimension for the images of those textures. As can be seen, there is little difference in the dimension values to allow segmentation or description. Figure 6(b)-6(d) shows segmentation results using dimension and lacunarity. Again, our new estimate of box dimensions, coupled with our lacunarity estimates produced excellent results.

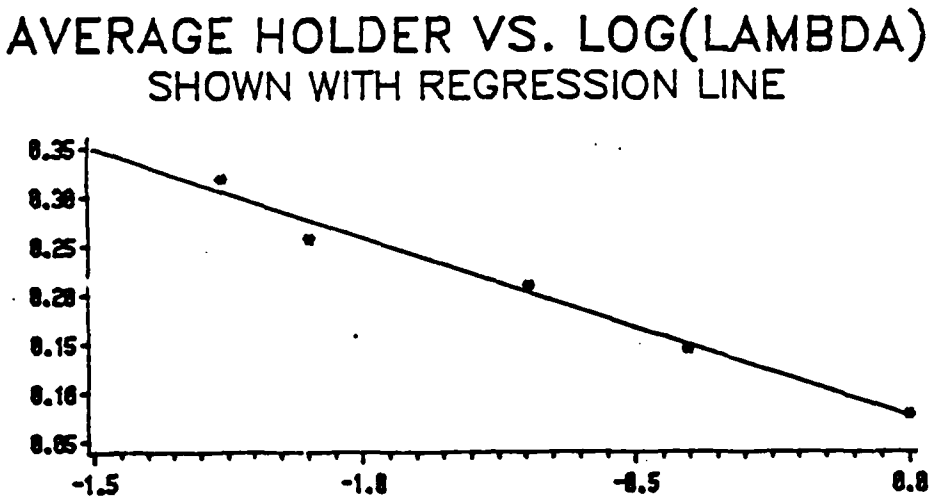


Figure 1. Plot of average Holder constant verses the log of scale factor.

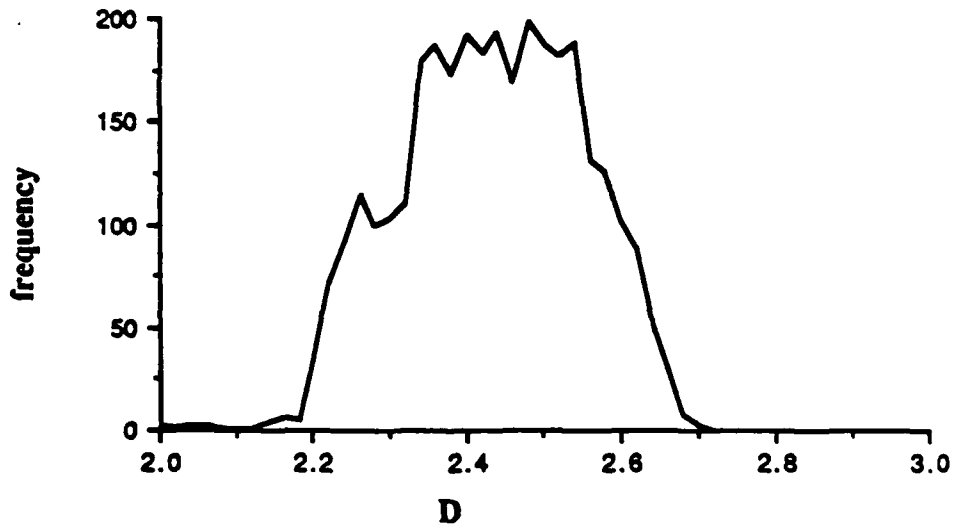


Figure 2. Histogram of estimated fractal dimension for windows in fractal mosaic by method in [9].

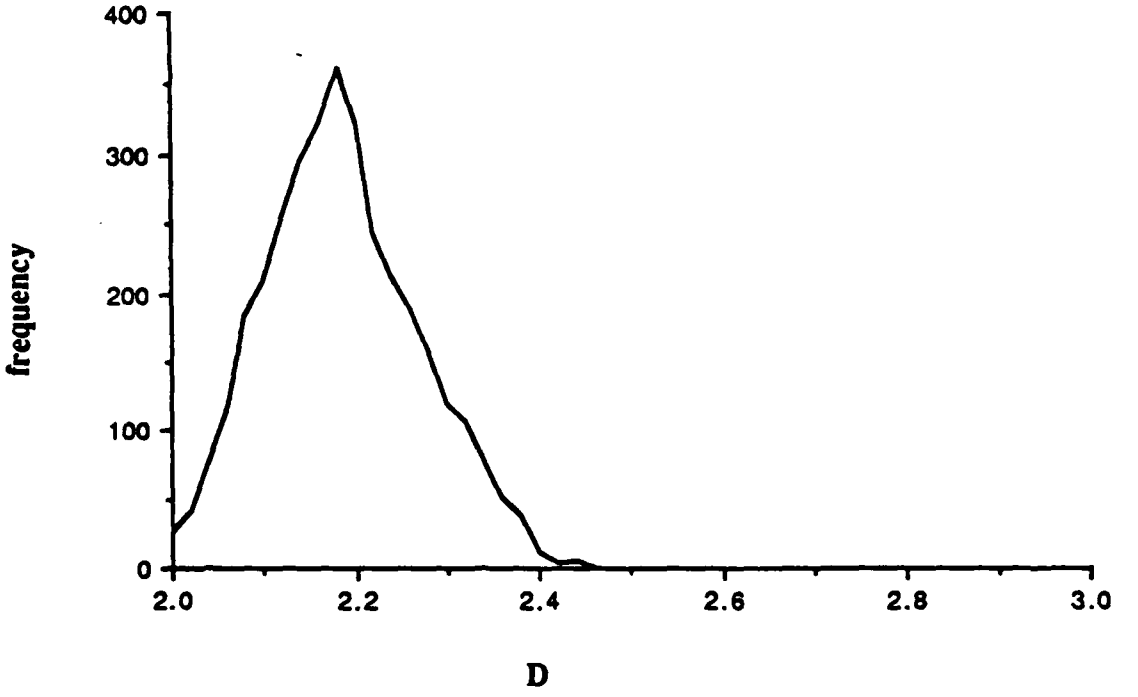


Figure 3. Histogram of box dimension estimated from fractal mosaic (non-interpolated).

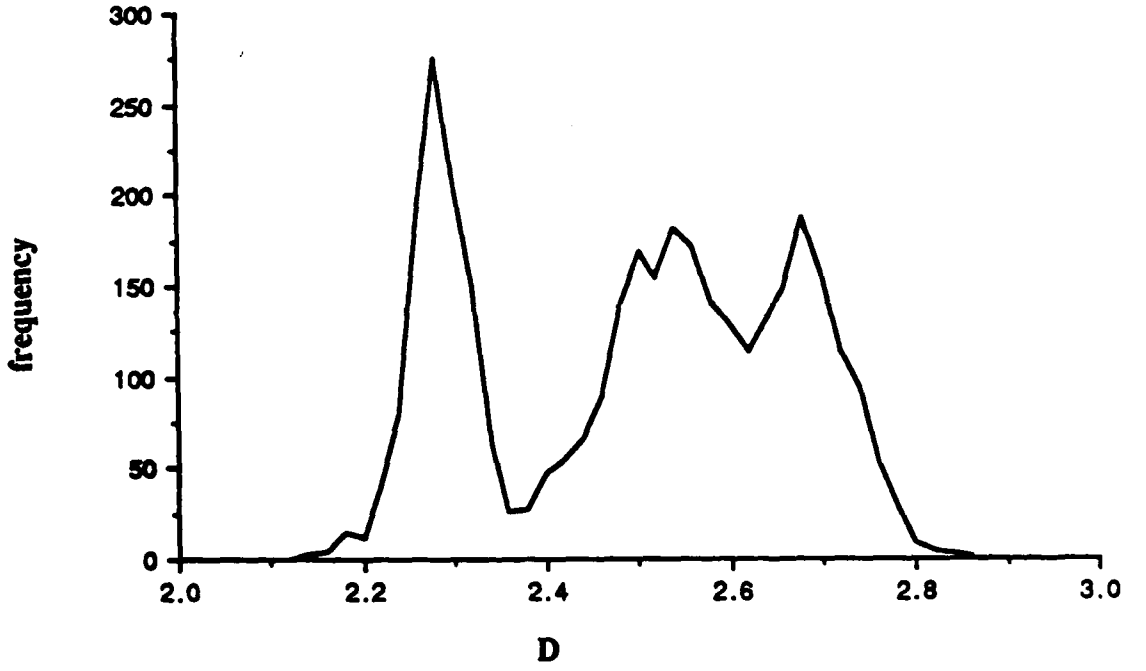


Figure 4. Histogram of box dimension estimated from fractal mosaic (interpolated)

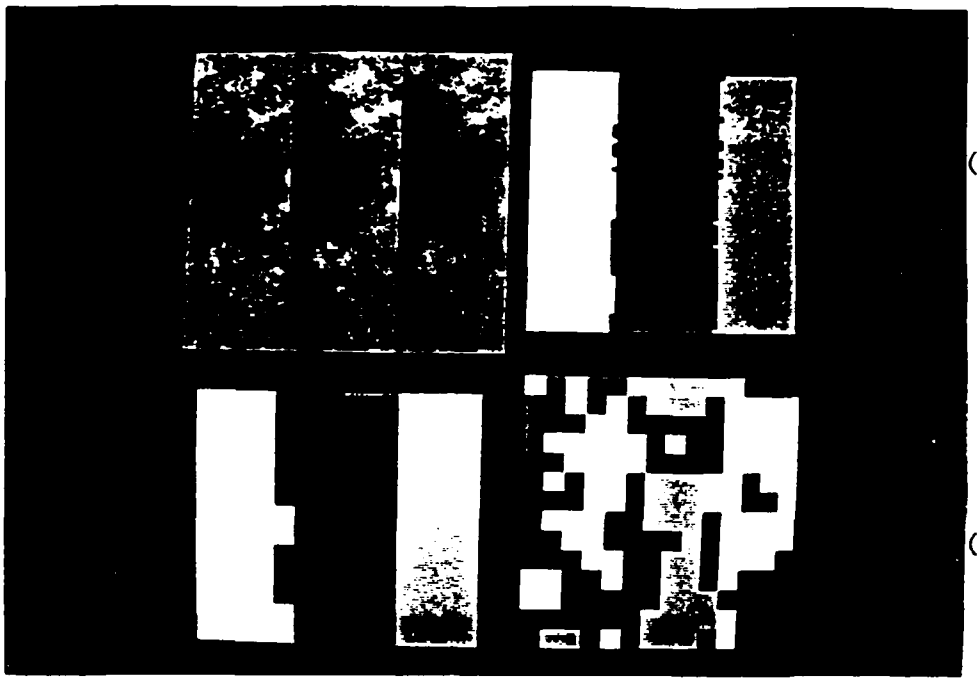


Figure 5. (a) Composite of three artificial fractal surfaces; (b) Segmentation using interpolated box dimension and lacunarity; (c) Segmentation using interpolated box dimension only; (d) Segmentation using non-interpolated box dimension only.

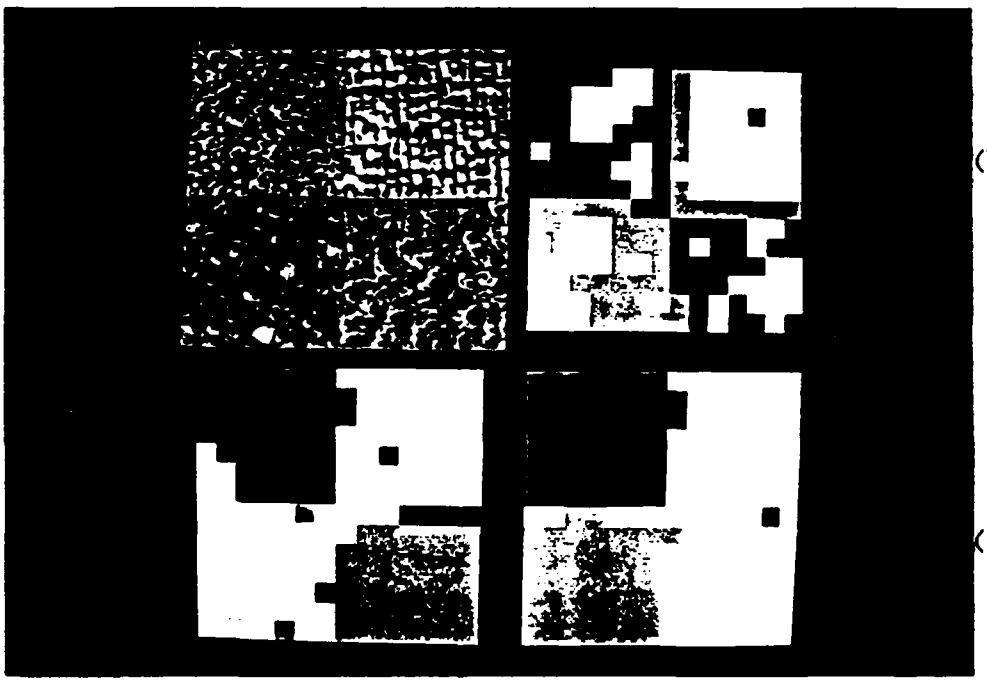


Figure 6. (a) Four texture composite; (b) Segmentation using 4 lacunarity features of Voss [13]; (c) Segmentation using 4 new lacunarity features and non-interpolated box dimension; (d) Segmentation using 4 new lacunarity features and interpolated box dimension.

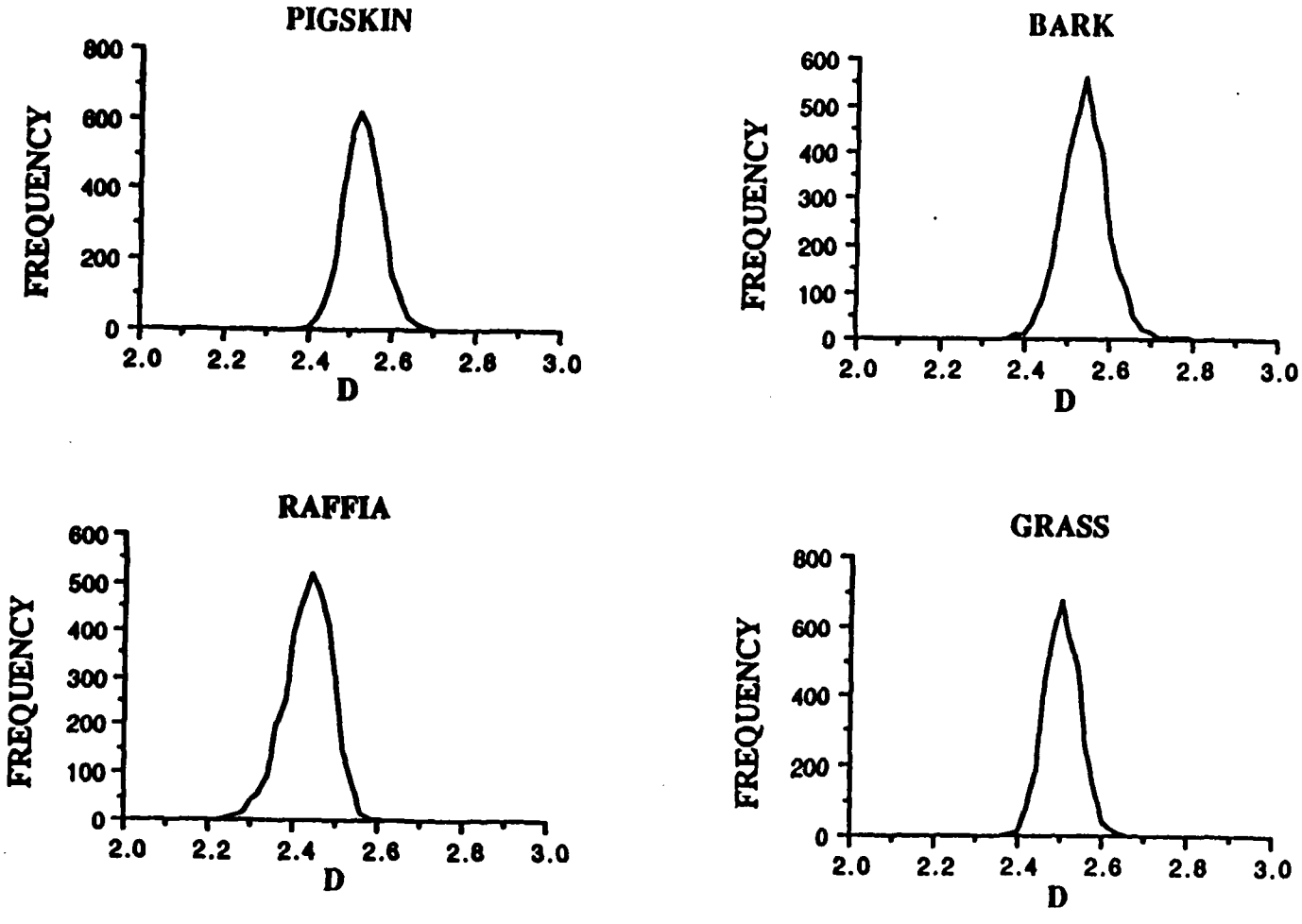


Figure 7: Histograms of estimated fractal dimensions for four textures.

REFERENCES

1. J. M. Keller, R. M. Crownover, and R. Y. Chen, "Characteristics of Natural Scenes Related to the Fractal Dimension", IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 9, No. 5, 1987, pp. 621-627.

2. S. Chen, J. Keller, and R. Crownover, "On the Calculation of Fractal Features from Images", 2nd Int. Conf. on Computer Vision, Tarpon Springs, FL, Dec. 1988, under review.

3. J. Keller, S. Chen, and R. M. Crownover, "Texture Description Through Fractal Geometry", Computer Vision of Graphics and Image Processing, accepted for publication.

4. S. S. Chen, "Fractal Geometry in Image Understanding", Ph.D. Dissertation, University of Missouri-Columbia.

5. B. Mandelbrot, The Fractal Geometry of Nature, Freeman, 1977.

6. B. Mandelbrot and J. N. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications", SIAM Review, Vol. 10, No. 4, 1968.

7. A. Pentland, "Fractal Based Description of Natural Scenes", IEEE Trans Pattern Analysis and Machine Intelligence, Vol. 6, 1984, pp. 661-674.

8. A. Pentland, "Fractal Texture", Proc. Int. Joint Conf. on Art. Int., 1983.

9. G. Medioni and Y. Yasumoto, "A Note on Using the Fractal Dimension for Segmentation", IEEE Computer Vision Workshop, Annapolis, MD, 1984, pp. 25-30.

10. S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple Resolution Texture Analysis and Classification", IEEE Trans. Pattern and Machine Intell., Vol. 6, No. 4, 1984.

11. T. Lundahl et al., "Fractional Brownian Motion: A Maximum Likelihood Estimator and Its Application to Image Texture:, IEEE Trans. on Medical Images, Vol. 5, No. 3, September, 1986.

12. Y. Ohta, K. Maenobu and T. Sakai, "Obtaining Surface Orientation from Texels Under Perspective Projection", 7th International Joint Conference on Artificial Intelligence, Canada, 1981.

13. R. F. Voss, "Random Fractals: Characterization and Measurement", Scaling Phenomena in Disordered Systems, Plenum Press, 1985.

5. A Least Squares Approach to Linear Discriminant Analysis

An important technique for object recognition and classification in image analysis, speech recognition, and other situations where intelligence is gleaned from comparison with training data, is to first reduce the dimension of the space in which the data is stored. Object data is represented as vectors is a high dimensional space, perhaps of dimension in the hundreds or thousands. These vector, along with the ones to be recognized, are projected into a very low dimensional space, even as low as one dimension, in such a way that the separation of prototypical classes is best preserved. It is then anticipated that object recognition algorithms will perform at their best because of the reduced computational load and reduced accumulation of round-off error.

Perhaps the best known of these dimension reduction techniques is Fisher's linear discriminant method [1-3,5,7]. Fisher's method requires the solution of a generalized eigensystem of the type also known as a generalized singular value decomposition. Early implementations called for computing large cross-product matrices, causing them to be ill-conditioned. They were also computationally expensive. Newer methods [8,9,11] improve that situation but don't take advantage of the rich structure inherent in the Fisher problem.

We present a new approach that takes advantage of the special structure of Fisher's method, is well-conditioned, and reduces the computational load. The idea is to project, as best possible in the least squares sense, the vectors in a given class onto a single point in the smaller dimensional space. First, a preliminary optimization problem is solved to find the right projection points. We have

developed two algorithms. The first, which is more stable in the nearly rank deficient case, uses the singular value decomposition. The second and faster method makes use of orthogonal triangularization (Q-R factorization). It is the method that should be used if updating the data vectors is anticipated. Both methods avoid potentially disastrous errors from calculating large cross-product matrices.

Suppose we have k vectors in R^n divided into c classes or clusters, with k_1 vectors in class 1, k_2 vectors in class 2 and so on through class c, having k_c vectors. Thus $k = k_1 + \ldots + k_c$. Let Z_i be the $n \times k_i$ matrix having columns that are the vectors in class i and let $Z = [Z_1, \ldots, Z_c]$ be the $n \times k$ matrix of all data vectors. We assume that Z is of full rank.

The problem we address is that of projecting the given data from R^n into a smaller dimensional space R^p ($p < n$; usually $p = c-1$) in such a way that correct identification of class membership can be determined from analysis performed in the smaller dimensional space. If the projection vectors are denoted $\phi_1, \ldots, \phi_p$ and if $\Phi = [\phi_1, \ldots, \phi_p]$, then the mapping to R^p is defined by $z_j \mapsto \Phi^T z_j$, $j=1,\ldots,k$.

In order to get the best separation in the projection space it is appropriate to use certain optimization criteria. The total scatter of a set of vectors $(z_1, \ldots, z_k)$ in R^n having mean m is

$$\text{total scatter} = \sum_{j=1}^{k} |z_j - m|_2^2.$$

Let $Y = [z_1 - m, \ldots, z_k - m]$, in terms of which

$$\text{total scatter} = \text{trace}(S_T),$$

where $S_T = YY^T$. The matrix S_T is called the total scatter matrix.

After performing the projection into a p-dimensional space (p is not a priori c - 1), the reduced total scatter becomes

$$\text{reduced total scatter} = \text{trace}(\Phi^T YY^T \Phi) = \sum_{i=1}^{p} \phi_i^T (YY^T)\phi_i = \sum_{i=1}^{p} \phi_i^T S_T \phi_i.$$

There are two other scatter matrices of interest to us, the within-class scatter matrix S_w and the between-class scatter matrix S_B. The matrix S_w is defined by

$$S_w = S_{w_1} + \ldots + S_{wc},$$

where for each i=1,...,c, S_{wi} is the total scatter matrix for class i alone. If X_i is the adjusted data matrix for class i, then $S_{w_i} = X_i X_i^T$. Letting $X=[X_1,\ldots,X_c]$, we have $S_w = XX^T$. The matrix S_B is defined by

$$S_B = \sum_{i=1}^{c} k_i (m_i - m)(m_i - m)^T.$$

Define

$$E = [E_1,\ldots,E_c] = \begin{matrix} k_1 \\ k_2 \\ \\ \\ k_c \end{matrix} \begin{pmatrix} \bar{1} & \bar{0} & \ldots & \bar{0} \\ \bar{0} & \bar{1} & \ldots & \bar{0} \\ & & \ddots & \\ \bar{0} & \bar{0} & \ldots & \bar{1} \end{pmatrix},$$

where $\bar{1} = [1,\ldots,1]^T$ and $\bar{0} = [0,\ldots,0]^T$. Let $M_c = [m_1 - m,\ldots,m_c - m]$. It is relatively easy to see that

$$Y = X + M_c E^T,$$

$$S_B = M_c E^T E M_c^T,$$

$$S^T = S_B + S_w, \text{ and}$$

$$YE = M_c E^T E.$$

An important tool in our analysis is the singular value decomposition (SVD) of the matrix Y [6,10]. The form of this decomposition that we use is written as

$$Y = U\Sigma V^T,$$

where in terms of the rank r of Y,

U is n x r and has orthonormal columns,

$\Sigma = \text{diag}(\sigma_1,\ldots,\sigma_r)$, where $\sigma_1 \geq \ldots \geq \sigma_r > 0$ are the nonzero singular values, and

V is k x r and has orthonormal columns.

Under the assumption that Z has full rank, $r = \min(n,k-1)$.

The generalized inverse of Y [6,10], denoted $Y^\dagger$, is given by

$$Y^\dagger = V\Sigma^{-1}U^T.$$

The criterion we impose is to maximize the reduced total scatter $\sum_{i=1}^{p} \phi_i^T S_T \phi_i$ subject to the constraints that $\phi_i^T S_w \phi_i = 1$, $i=1,\ldots,p$. The constraints say that the projections of the within-class scatter in each coordinate are bounded by one. Since $S_T = S_B + S_w$,

$$\sum_{i=1}^{p} \phi_i^T S_T \phi_i - \sum_{i=1}^{p} \phi_i^T S_B \phi_i + \sum_{i=1}^{p} \phi_i^T S_w \phi_i$$

$$- \sum_{i=1}^{p} \phi_i^T S_B \phi_i + p.$$

Thus an equivalent formulation of the problem is to maximize the reduced

between-class scatter $\sum_{i=1}^{p} \phi_i^T S_B \phi_i$ subject to the same constraints. This is the Fisher linear discriminant problem [2,3].

The Lagrange function for the Fisher problem is

$$F(\phi_1,\ldots,\phi_p,\mu_1,\ldots,\mu_p) = \sum_{i=1}^{p} [\phi_i^T S_B \phi_i - \mu_i(\phi_i^T S_w \phi_i - 1)].$$

Setting $\text{Grad}_{\phi_i}(F)=0$ gives

$$S_B\phi_i = \mu_i S_w \phi_i, \quad i=1,\ldots,p, \qquad (*)$$

which is a generalized eigensystem.

The analysis splits into two cases depending on the rank of X, which is min(n,k-c):

i) the overdetermined case (which we take to include the exactly determined case), in which $k-c \geq n$ and

ii) the underdetermined case, in which $k-c<n$.

The overdetermined case is the one that has received wide attention and is best understood. In this case the rank of S_w is n, which implies that the generalized eigensystem (*) is equivalent (analytically, but not for computational purposes) to the ordinary eigensystem $S_w^{-1} S_B \phi_i = \mu_i \phi_i$, $i=1,\ldots,p$. The matrix S_B has rank c-1, from which it

follows that the appropriate value of p, the dimension of the reduced space, is c-1.

It is inadvisable to compute the scatter matrices S_B and S_w as steps in solving the eigensystem (*). The computation of $S_w = XX^T$, in particular, can be unstable and even if computed exactly, its condition number $\kappa(XX^T)$ can be large since $\kappa(XX^T) = \kappa(X)^2$. This can result in large relative output errors when solving the generalized eigensystem.

A stable computation of solutions of (*), or equivalently of

$$(M_c E^T)(M_c E^T)^T \phi_i = \mu_i XX^T \phi_i, \quad i=1,\ldots,c-1,$$

can be carried out using one of the recent treatments of the generalized singular value decomposition [8,9,11] However, these methods do not take advantage of the special structure of the Fisher problem, as our methods do, and are not particularly amenable to updating.

The following theorem is the key to our proposed methods. It can be interpreted as saying that we can solve the Fisher problem by mapping as best possible in the least squares sense all the vectors in a given class onto a certain vector in the projection space. It also describes how to find these critical vectors.

Theorem 1. Let $\phi_1,\ldots,\phi_{c-1}$ be the projection vectors for the overdetermined case. Then for each i=1,...,c-1, there exists a vector $x_i \in R^c$ such that $\phi_i = Y^{\dagger T} E x_i$. Moreover, $x_1,\ldots,x_{c-1}$ are the eigenvectors of the c x c generalized eigensystem $E^T VV^T Ex = \lambda E^T Ex$ that are different from $x = [1\ldots1]^T$.

(The proof of this and other theorems in our development may be seen in the full paper, which is included in the appendix.)

Theorem 2. If c=2 in the previous theorem so that there is just one x_i (call it x), then up to scalar multiples it can be assumed to be given by $x=(-k_2 k_1)^T$.

The underdetermined case is characterized by the condition that S_w has rank less than n, or equivalently, that $k<n+c$. In this situation every complex number is a generalized eigenvalue for (11) corresponding to the generalized eigenvector $\phi=[1,\ldots,1]^T$.

However, there is a different criterion that is appropriate for the underdetermined case and for which there is a solution to the resulting optimization problem [1]. The new criterion is to maximize the reduced total scatter $\sum_{i=1}^{p} \phi_i^T S_T \phi_i$ subject to $\phi_i^T S_w \phi_i=0$, $\phi_i^T \phi_i=1, i=1,\ldots,p$.

Theorem 3. In the underdetermined case, the problem of maximizing the reduced total scatter subject to the constraints $\phi_i^T S_w \phi_i=1$ is equivalent to

1) solving the generalized eigensystem

$$(E^TE)x_i=\lambda_i A^T A x_i, \quad i=1,\ldots,c-1,$$

where $A=\Sigma^{-1}V^TE$ and $\lambda_1,\ldots\lambda_{c-1}$ are nonzero,

2) setting $\phi_i=(Y^\dagger)^T E x_i$.

3) normalizing ϕ_i, $i=1,\ldots,c-1$.

The conclusion of Theorem 2, that up to constant multiples, $x=(-k_2 k_1)^T$, also holds for the underdetermined case.

Theorem 4. Let Φ be the output from the algorithm for the underdetermined problem. Then Φ has orthogonal columns.

For a faster algorithm in the overdetermined case, at the possible expense of losing some stability in the nearly rank deficient case, it is better to use the Q-R factorization of Y^T as a tool. It is also the preferred method if updating the projection vectors is anticipated. In the overdetermined case, Y^T is a $k\times n$ matrix with $k\geq(n+c)$ and $\text{rank}(Y^T)=n$. The Q-R factorization produces a $k\times n$ factor Q having orthonormal columns

and an nxn nonsingular upper triangular factor R for which $Y^T=QR$.

Theorem 5. The projection vectors for the overdetermined case can be calculated from $\phi_i=Y^{\dagger T}Ex_i$, i=1,...,c-1, where $x_1,\ldots,x_{c-1}$ satisfy

$$M_c^T Y^{\dagger T} E x_i = \lambda_i x_i$$

for nonzero scalars $\lambda_1,\ldots,\lambda_{c-1}$.

Algorithm 1. SVD version. Overdetermined and underdetermined cases.

1) Calculate Y.
2) Calculate the SVD(Y) = $U\Sigma V^T$.
3) If the problem is overdetermined calculate $A = V^TE$. If underdetermined, calculate $A = \Sigma^{-1}V^TE$.
4) Calculate the right singular vectors corresponding to the nonzero singular values of AG^{-1}, where $G = \sqrt{E^TE}$ (= $\mathrm{diag}(\sqrt{k_1},\ldots,\sqrt{k_c})$). Let ψ be the cx(c-1) matrix whose columns are these vectors.
5) Calculate $X=G^{-1}\psi$.
6) If the problem is overdetermined, calculate $\Phi=U\Sigma^{-1}AX$. If underdetermined, calculate $\Phi=UAX$.
7) Normalize ϕ_i, i=1,...,c-1, if desired.

The main computational load in this algorithm is step 2, the singular value decomposition, which requires about $7kn^2 + (11/3)k^3$ flops in the overdetermined case.

Special note for the case c=2.

If c=2, omit steps 3, 4, and 5, for in this case there is just one x_i (namely x_1) and up to scalar multiples, it can be taken to be $x_1=(-k_2k_1)^T$.

Algorithm 2. Q-R version. Overdetermined case only.

1) Calculate Y.

2) Calculate the Q-R factorization $Y^T=QR$, where Q is kxn and has orthonormal columns, R is nxn, upper triangular, and nonsingular.

3) Calculate $F=[F_1,\ldots,F_c]$ by

   a) finding $b_j=Q^TE_j$, $j=1,\ldots,c$, where $E = [E_1,\ldots,E_c]$, and

   b) solving $RF=B(=[b_1,\ldots,b_c])$ for F.

4) Calculate $A=M_c^TF$.

5) Calculate the eigenvectors $x_1,\ldots,x_{c-1}$ of A corresponding to nonzero eigenvalues.

6) Calculate $\phi_j=Fx_j$, $j=1,\ldots,c-1$.

7) Normalize ϕ_i, $i=1,\ldots,c-1$, if desired.

The main computational load in this algorithm is step 2, the Q-R factorization of Y^T. It requires about $n^2(k-n/3)$ flops if done by Householder transformations.

Special note for the case c=2.

If c=2, omit steps 4 and 5, for in this case there is just one x_i (namely x_1) and up to scalar multiples, it is given by $x_1=(-k_2k_1)^T$.

It may be desirable to append or delete a class of data vectors to the linear discriminant problem after the original calculations have been done and to quickly update the projection vectors. Techniques have been known for some years for updating a Q-R factorization [4,6] and we have drawn on these methods for updating the projections vectors. The process of updating a singular value decomposition is not as easy and hence we have restricted attention to updating in the overdetermined

case in which the Q-R factorization is used (Algorithm 2). Our presentation concludes with an analysis of the problem of appending a new data class. The dominant term in an estimate for the computational load for updating is $n^2 k_{c+1}$ flops, where k_{c+1} is the number of vectors in the appended class. This compares favorably with the estimated $n^2(k+k_{c+1}-n/3)$ flops estimated for starting over when new data is appended.

REFERENCES

1. R. M. Crownover and J. M. Keller, "Fast dimension reduction that preserves underdetermined data clusters", Proc. SPIE Conf. Advanced Signal Processing Algorithms and Architectures, San Diego, (1986), pp. 238-244.

2. R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, John Wiley & Sons, New York, 1973.

3. R. A. Fisher,"The use of multiple measures in taxonomic problems", Ann. Eugenica, 7, Part II (1936), pp. 179-188.

4. P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, "Methods for modifying matrix factorizations", Math. Comp. 28 (1974), pp. 505-535.

5. G. Hobson, J. Keller, and R. Crownover, "Method of extracting uncorrelated features from gray level images for target recognition", Proc. Eight Ann. Symposium on Ground Vehicle Signatures, Houghton, MI, (1986), pp. 208-214.

6. G. H. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins Univ. Press, Baltimore, 1983.

7. J. M. Keller, R. M. Crownover, B. Hovson and J. Wootton, "Target recognition using the Karhunen-Loeve transform", Proc. Int. Conf. Systems, Man and Cybernetics, Tucson, (1985), pp. 310-314.

8. C. C. Paige, "Computing the generalized singular value decomposition", SIAM J. Sc. Statist. Comput., 7 (1986),pp.1122-1146.

9. G. W. Stewart, "Computing the CS decomposition of a partitioned orthonormal matrix", Numer. Math., 43 (1982), pp. 297-306.

10. G. W. Stewart, Introduction to Matrix Computations, Academic Press, New York, 1973.

11. C. F. Van Loan, "Computing the CS and generalized singular value decompositions", Numer. Math., 46 (1985), pp. 479-491.

6.0 Rule-Based Automatic Target Recognition

Introduction

The majority of automatic target recognizers undertaking field evaluation today owe their internal structure to a classical statistical approach. Although the dimensionality of the variable parameters that each system is subject to is large, little use is made of context and ancillary information such as time of day, sensor, weather conditions and intelligence data. Such ancillary data can be profitably used to alleviate the algorithmic burden of accommodating the extreme ranges of conditions.

Presented in this section are two novel approaches to include ancillary knowledge into the control structure of an automatic target recognizer (ATR). Automatic target recognition involves the determination of objects in natural scenes in different weather conditions and in the presence of both active and passive countermeasures and battlefield contaminants. This high degree of variability requires a flexible system control capable of adapting as the conditions change. The desired flexibility can be achieved with a rule-based system in which the knowledge of the effects of scene content and ancillary information on algorithm choices and parameter values can be modelled and manipulated. This on-going effort is an outgrowth of earlier activity in automatic target recognition research. New theoretical and practical tools were developed for the analysis of military scenes, with emphasis placed on methods to deal with uncertainties associated with such imagery [1-5]. The ongoing effort involves incorporating the knowledge and experience gained in working

with such imagery and with modelling uncertainties into a rule-based structure for the detection and recognition of objects in military scenes.

A major focus of the research as reported here for incorporating AI into automatic target recognition has been the use of context to cue the possible likelihood of a target in a given area of the scene [6]-[11].

6.1 Numerical Uncertainty Propagation System

The approach reported on here was borne out of several years of independent research in image processing, image analysis, image understanding and artificial intelligence techniques. Because of the large variability in automatic target recognition, no single set of algorithms, no matter how adaptive they could be made, would give consistent, reliable results when subject to the full variety of target conditions and scenario conditions. Yet, it was realized that by having knowledge of the conditions (that could be measured by simple metrics) that an expert analyst could select an appropriate algorithm which could yield an optimal performance.

The known tool for implementing this expert corporate knowledge is a rule based system. It is desired that such a system indicate when it was being subjected to a situation for which there was no supporting research. It is desired that the system be structured so that it identified circumstances which were outside its experience. Finally, it is not necessary to have a set structure for finding targets and target types but to let the data trigger the rules for finding "potential objects". Once a potential object has been found the system carries

multiple hypotheses as regard to the type of the object using a local Dempster-Shafer approach to eventually determine the target type.

Classical statistical image processing generally follows the stages of enhancement, prescreening, segmentation and feature extraction followed by detection, recognition and/or identification. Within each stage, the image analysts developed a whole series of algorithms which themselves had adaptive coefficients, or thresholds [12, 13]. Indeed, some of these algorithms themselves included simple rule-based algorithms.

The systems usually utilize deterministic and/or statistical rules for classification. Little use is made of context and ancillary information such as time of day, season, weather conditions and "intelligence" data. The conventional approach utilizes a training data set as a basis to select processing algorithms, select parameter values, select feature sets and to build decision rules. If an actual situation fell outside of the training set, such a system would make a decision with a relatively high, and likely unacceptable, error probability.

In addition to developing individual algorithms, measures for evaluating the performance of an algorithm under various circumstances were also developed. These "tractability" fundamental parameters are i) size, ii) contrast, iii) clutter, iv) motion, v) shape and vi) color. The real worth of this work was that an image understanding analyst can quantify image tractability as a function of image metrics [14,15]. A set of algorithms could, in general, be selected and their coefficients controlled to give very reasonable performance provided that the set of test images to which the system was subjected was confined to small excursions of the image conditions.

Combining knowledge about image processing with that about scene understanding became the goal of the artificial intelligence task. At an early stage, the knowledge was classified under the general headings of sensor, scene, and objective. The impact of the sensor is self evident. The significance of the scene can be illustrated by comparing the difficulty of finding a man made object in a rural scene as distinct from an urban scene.

The target condition itself plays a dominant role. Apart form its size, its signature against its background is one of the most dominant video features. Clearly, a well camouflaged target will present little or no contrast to the ATR which makes the basic task of initial target detection difficult [15]. Conversely a sharply contrasted target is very easy to locate. The weather conditions greatly contribute to this received image contrast. Rain, fog, and BIC (battlefield induced contaminants) all contribute to image degradation.

The final impact upon an ATR structure is the mission objective. In the first level of classification under the objective category are detection, recognition and identification. If all that is required is detection then essentially the ATR is faced with a two class problem of target or non-target. Once the target category has been designated as a military set (tanks, APCs, trucks, helicopters, etc.) then every object not recognized as one of that target set is a non-target. If recognition is the requirement, then the ATR has to accommodate a multiclass structure and look for a set of features which can discriminate between those classes selected. Identification of a particular object subclass leads to yet a more tasking problem In addition to the baseline objective of detection, recognition and

identification, the target type and its priority impact the algorithms chosen. For Ground order of battle, targets of opportunity, such as tanks, etc., tend to proliferate in the battlefield. An ATR carried on a high valued asset such as a fighter aircraft, whose objective is to find such targets of opportunity should be optimized to have a very low false alarm rate at the expense of probability of detection. On the other extreme are high value targets such as a mobile nuclear missile site for which one should be prepared to accept a higher probability of detection at the expense of a higher false alarm. Again, ancillary data on the target can greatly ease the target detection and recognition problem. For instance the majority of high value targets such as bridges, POL dumps, etc. have known physical locations. The linking of the knowledge concerning the location of the sensor platform (-aircraft-) and look angle of the sensor can indicate to the ATR that the object should be there geographically. The ATR in this case has much of the responsibility of finding the target removed by the inclusion of inertial navigation data and the knowledge that the target is at that location.

This knowledge was incorporated into an ATR control structure through the use of a rule-based network, or tree, to determine i) the choice of processing algorithms, ii) the order of application of these algorithms, iii) the parameters utilized in these techniques and iv) provide an overall confidence level associated with the final decision. Such a rule-based structure offers the following advantages in ATR applications:

1) efficient use of knowledge directly concerned with relationships of conditions to conclusions or deductions, such as the influence of ancillary factors on selecting features or the image sensor;
2) isolation of (IF...THEN rules) for feature selection from those for selecting a segmentation algorithm;
3) ability to handle uncertainties in terms of probabilities, belief functions or possibility distributions;
4) ability to perform specific goal directed hypothesis testing, for example, different approaches are necessary to decide if an object is a tank, or more generally a target.

Description of the Process.

Early image analysis research had followed a sequential approach. It was found that one could perform considerable enhancement either globally or locally, which would give "pleasing" results to a human observer [16]. However, such preprocessing did little in terms of improving an ATRs overall performance unless it was concerned with removing an artifact generated by the sensor.

Similarly, many of the common prescreeners were studied and tested using different sensors against targets in different scenarios [17]. What was learned was that, depending on the sensor and the scenario, not only is there an appropriate choice of prescreener but there is also an optimal choice of that prescreener's coefficient [17-18].

Therefore it became evident that one could define a knowledge base structure such that if the sensor was known, and the scene conditions measured, the most appropriate choice of preprocessor and prescreener could then be selected.

The next stage became somewhat more complex. Research had been performed on segmentors [14], features extraction [17] and the methods for combining features [17, 18] to determine the detection, recognition or identification of objects in military scenes. In an effort to optimize the segmentation process it was realized that the measure of performance of the segmentor was related to the consistency of the feature it produced for the feature extraction stage [14]. The value of the feature itself depended upon the manner in which that feature separated the chosen class from the other classes. Through considerable testing of a number segmentors (around 20) in conjunction with over 150 features there evolved a consistent set of appropriate feature extractor-segmentor pairs that gave reliable results [14, 17].

The appropriate set of features themselves was predicted upon the class of targets required (i.e. the objective). For each classification problem encountered, target vs. non-target, tank vs. APC, helicopter vs. false alarm, etc., the effects of different collections of features and pattern recognizers on a large data base of military objects were studied. The classifiers included Bayes decision rule, crisp and fuzzy K nearest neighbor and perceptron schemes, and a Dempster-Shafer evidence combination process [4, 5, 18-21]. Then, rules were developed which paired the decision making procedure with the appropriate feature sets for each subproblem. This in turn demanded an appropriate choice of feature-extractor methodology.

The work in determining the optimal classifiers was based upon extracting the features from the training data into known ranges of value. A dilemma was that when features were extracted from test data that did not correspond to the range of data extracted from the training

set, then any decision derived from that data is unsubstantiated. To accommodate this, the ATR was allowed to address other features which might fall into a range that had been observed from the training data. Should the ATR not find a feature that corresponds to the test data range, it returns an answer of "outside our experience."

As an added robustness feature, multiple hypotheses as to the identity of the potential target were carried. For instance, in determining whether a detection from the prescreener is a target (defined as a tank or an APC) or a false alarm, the possibility that it was a target or a non-target, was examined, or that it was an APC or a non-APC, or that it was a tank or non-tank. Whatever the objective of system at a particular time, the first piece of evidence acquired is a detector decision (target vs. non-target) using a Bayes Rule with 8 features chosen from the feature set. The rule was trained on approximately 900 targets and around 100 false alarms. The decision made at this stage (together with the Bayes confidence) is used to tailor the resulting evidence acquisition.

From this point on, different pattern recognition problems are solved, and the results combined in a voting scheme individualized to the overall objective of the system. For example, if the objective were to distinguish tanks, APC's and false alarms, then the following subproblems were initiated to provide the evidence for the final vote: tank vs. non-tank, APC vs. non-APCs, targets vs. false alarms, APCs vs. false alarms, and tanks vs. false alarms. In each of these processes, four features chosen from the feature set as being good separators of the training data were used. The evidence combination in each decision process was based on Dempster-Shafer belief theory [21]. For each

feature, a simple support function was generated separately for each hypothesis. It is in this support function that information "outside our experience" can be ignored. In fact, the basic probability assignment for the hypothesis under consideration was calculated by a π-function centered on the interval of the feature axis occupied by the training data. Hence, if this feature value for a test object does not fall into the range of our training data, a vacuous support function is generated. These support functions are combined using Dempster's Rule into a belief function related to this feature. These belief functions are then combined for all features to produce overall beliefs for the hypotheses under consideration. A decision is made to favor the hypothesis with greatest belief. This structure was chosen so that spurious values of a few features caused by noise, partial occlusion, etc. will not overly bias the decision, as may happen in a Bayes technique.

There are several rules in the system which combine the decisions generated by those subproblems into a final classification. These rules depend upon the overall objective, and the choice and results of the various subtests. Intuitively, if there is a clear winner with high enough belief, then that hypothesis is chosen. However, there are tie-breaking rules which reflect mission objectives; for example, if the object could be either a tank or an APC with about equal support, then call it a tank. There are even rules which override a majority if the evidence from a high priority rule is strong enough. An instance of this type of rule occurs where the object is thought to be a target from Bayes rule, and a tank in the tank vs. non-tank rule (with high enough belief) but is labeled a false alarm by two or three rules which allow

false alarm as a hypothesis. In this case, the majority is overridden by the confidence in the target and tank decisions.

Implementation and Results

The rule based system control strategy contained both forward chaining and backward chaining paradigms. The forward chaining approach tends to follow the traditional sequence of steps in image analysis, while the backward chaining mode allows specific tailoring of the system structure for hypothesis generation and testing.

The system was implemented in an expert system shell for rapid prototyping with image inputs from the Computer Vision Laboratory equipment at UMC. The basic prototype system uses 212 rules.

The object recognition portion of the rule base was tested on a sequence of 100 frames containing two tanks and an APC during which the APC moves behind one of the tanks and into and out of a ravine. This sequence of images is considerably different from the data used to "train" the rule base. Using the output of the first part of the rule base, the images were prescreened and segmented and the chosen features were extracted. The rule base was executed using different objectives and representative results are displayed in Table 1. The format of the table gives the actual object under consideration, the system objective, the result of the target detection stage and the recognition evidence, followed by the final classification. The local Shafer belief values for individual processes have been suppressed and here only the partial determination are reported. It can be seen from the table that different system objectives give rise to different recognition processes and interpretation rules. Several instances of system behavior can be

highlighted. There are circumstances when this rule base will not make a partial decision. This occurred when the results of the individual recognition procedures conflicted with the system objective. This can be seen in tests (4), (5), (8) and (10) in Table 1. In each case, an Undecided (Nil) response was generated, and the search for evidence proceeded. With the object labeled APC 1 (8), in the APC vs. tank scenario, the tests tank vs. non-tank and APC vs. false alarm indicate that the result is false alarm, contrary to the objective. The system recorded a Nil decision from the evidence, and proceeded. When it encountered the same confusion in the next set of tests, it produced a final classification of Undecided. (This APC was partially occluded by the ravine). However in the APC vs. false alarm case, a decision was made quickly (This APC's features resembled those of ground clutter due to its occlusion). Now APC 4 corresponds to the APC after it had become completely visible and was correctly classified in all scenarios.

Test (6) from Table 1 involved a tank in a three class recognition problem In this case, since false alarm was a viable answer, the inconsistency described above was not present. Hence, there were two votes for false alarm in this case, countered by the one vote for tank. However, the fact that the object was called a target by the detection algorithm (Bayes Decision Rule), and a tank in the tank vs. APC process, together with "high" belief in tank and "low" belief in false alarm, the system correctly identified the object. This is an example of a non-majority decision rule consistent with a mission plan.

The system described above represented the initial prototype ATR and is denoted as the "polling" system. Three enhancements were made to this system during the latter stages of the grant. The calculation of

the simple support functions was changed from a simple π-function to the fuzzy integral, as described in section 3.2. The same basic voting and polling structure was used. A comparison of these two systems can be found in Table 2. As can be seen, the fuzzy integral provided a better evaluation of feature evidence resulting in fewer misclassifications. Then for the three class problem (Tank vs APC vs False Alarm), the control structure of the system was modified. Instead of voting on the various subproblems, each such rule produced a belief function over the frame of hypotheses. These belief functions were then combined globally using Dempster's Rule to obtain a final classification (hypothesis with largest belief). The results for this approach using π-functions and the fuzzy intrgral within each rule are also displayed in Table 2. The π-functions proved to be too simplistic of an approach. When the polling strategy was removed, the overall classifications went down. However, better results were obtained from the fuzzy integral in this new structure.

Table 1. Sample Output*

Actual Object	System Objective	Evidence		Final Classification
		Detection	Recognition [(process)/result]	
Tank 11				
(1)	T vs FA	Target	(T vs A)/T (T vs NT)/T	Tank
(2)	T vs A	Target	(T vs A)/T (T vs NT)/T (A vs NA)/NA (T vs FA)/T	Tank
(3)	T vs A vs FA	Target	(T vs A)/T (T vs NT)/T (A vs NA)/NA (T vs FA)/T	Tank
Tank 13				
(4)	T vs A	Target	(T vs A)/T (T vs NT)/T (A vs NA)/NA Nil (T vs FA)/FA	Tank
Tank 22				
(5)	T vs A	Target	(T vs A)/T (T vs NT)/NT Nil (A vs FA)/FA (A vs NA)/NA Nil (T vs FA)/FA	Tank
(6)	T vs A vs FA	Target	(T vs A)/T (T vs NT)/NT FA (A vs FA)/FA (A vs NA)/NA FA (T vs FA)/FA	Tank
APC 1				
(7)	A va FA	False Alarm	(A vs NA)/NA	False Alarm

(Table 1 continued)

(8)	A vs T	False Alarm	(T vs NT)/NT	Nil	
			(A vs FA)/FA		
			(A vs NA)/NA	Nil	
			(T vs FA)/FA		Undecided
APC 4					
(9)	A vs FA	Target	(T vs A)/A		
			(A vs NA)/A		APC
(10	T vs A	Target	(T vs A)/A		
			(T vs NT)/NT	Nil	
			(A vs FA)/FA		
			(A vs NA)/A		APC
Clutter					
(11	T vs A vs FA	False Alarm	(T vs NT)/NT		
			(A vs FA)/FA		
			(A vs NA)/NA		False
			(T vs FA)/FA		Alarm
			(A vs FA)/FA		

* Abbreviations: T - Tank
NT - Not Tank
A - APC
NA - Non APC
FA - False Alarm

Table 2
Confusion Matrices for ATR Testing Results *

a) Polling using π-functions

	Tank	APC	FA
Tank	166	10	0
APC	20	10	36
FA	0	0	13

b) Polling using Fuzzy Integral

	Tank	APC	FA
Tank	139	37	0
APC	15	47	4
	0	0	13

(Table 2 continued)

c) D-S using π-functions

	Tank	APC	FA
Tank	148	9	19
APC	20	15	31
FA	0	0	13

d) D-S using Fuzzy Integral

	Tank	APC	FA
Tank	142	34	0
APC	13	53	0
	0	0	13

* Polling refers to local Dempster's Rule with a voting strategy. D-S refers to global use of Dempster's Rule across rules.

6.2 Fuzzy Logic Automatic Target Recognition System

We now describe a fuzzy rule-based production system which incorporates the complexity and uncertainty into the model and provides a natural language interface to both midlevel and the AI high-level vision subsystems. In effect we are relaxing the desire for perceived precision as found in numeric models in an effort to increase the significance or believability of the results. Understanding the contextual content of the image is an important feature of this rule-based system. The extension of this contextual knowledge to the fuzzy logic system is used to resolve conflicting interpretations or to refine initial analysis.

We have applied the fuzzy rule-based production system described in Section 3.1 to two areas of automatic target detection and recognition. The application involved the use of temporal sequences of forward-looking infrared (FLIR) and TV images. The system consists of three distinct processing phases: (1) prescreening, (2) scene recognition, (3) contextual knowledge-based validation. Each of these processes is described in detail in the following sections along with

some sample rules which highlight the various concepts. A full listing of the rules used in this experiment can be found in the appendix.

Prescreening

The first step of an automatic target recognizer is to prescreen the individual image frames by either a series of size-contrast or spoke filters to find regions containing possible objects of interest. Extracting these regions involves an exhaustive search, that is, the system needs to try many different prescreened windows. For our automatic target recognizer to be able to do the search in a short amount of time, some sort of task-dependent knowledge was introduced. For example, a typical rule is

(RULE 1) If:

the range is long

Then:

the prescreened window size is small.

The values of the linguistic variables for primary terms were modelled by trapezoidal numbers over the specified domains given by

$$\text{trap}(v;a,b,c,d) = \begin{cases} 0, & v<a \\ \dfrac{1}{b-a}(v-a), & a\le v\le b \\ 1, & b\le v\le c \\ \dfrac{1}{c-d}(v-d), & c\le v\le d \\ 0. & v\le d \end{cases} \qquad (1)$$

The hedges very and more or less are functional models of the primary terms as defined in [20]. For example, the values long and small were represented by

$$\text{long} = \text{trap}(u;700,900,1000,1000), \qquad (2)$$
$$\text{small} = \text{trap}(v;1,1,5,15),$$

where u is measured in meters and v is an integer value representing the size of the window.

In this experiment we computed the linguistic values for the range at any distance (in meters), denoted by D, as follows. Since there is a considerable uncertainty in the distance of the object to the sensor (because of the approximation of the range), the uncertainty inherent in this value for the data set available was modelled by trapezoidal number given by

$$\text{range} = \text{trap}(u;u1,u2,u3,u4), \qquad (3)$$

where

$$u1 = \max(0, D-200),$$
$$u2 = \max(0, D-100),$$
$$u3 = D+100,$$
$$u4 = D+200.$$

This value is then matched to the nearest (in Hamming distance) linguistic term in the database and the result is used as the input to the system.

One interesting note in the resulting window size obtained by the above rule is that an AI routine can be used to generate window sizes for the prescreener using right and left α-level set endpoints (rounded up to the nearest integer). For example, if the range is **long** then by using the above definition for the linguistic value **small**, the right α-level set endpoints generated at intervals of 0.2 from 0 to 1 are 15, 13, 11, 9, 7, and 5, and the left α-level set endpoints are all 1. These level set endpoints can be then translated into window sizes 15x15, 13x13, 11x11, 9x9, 7x7, and 5x5, and 1x1, respectively. On the other hand, if the range is more or less **long**, then the value of the linguistic window size will be more or less **small**. By generating the level set endpoints as before and after translation the respective window sizes are 15x15, 15x15 (rounded up), 14x14, 12x12, 9x9, 5x5, and 1x1. It can be seen that the latter window sizes are at least as large as the former ones. The AI routine can then make a decision about which one of the hits are likely to be the target based on the belief used in the value of α. i.e., the higher the α, the higher the confidence.

Scene Recognition

In scene recognition, for a particular scenario, we primarily worked on two types of objects: man-made and natural. In this experiment, man-made objects consisted of armored personal carrier (APC), and TANK. Natural regions consisted of ROAD, SKY, FIELD, TREE, etc. We modeled the values for a linguistic variable confidence by fuzzy sets over a common domain [0,1] which are set up to convey the meaning of natural language expressions. For example, we modeled our primary terms high, medium, and low by the following:

$$\begin{aligned} \text{low} &= \text{trap}(u;0,.1,.2), \\ \text{medium} &= \text{trap}(u;.4,.45,.55,.6), \\ \text{high} &= \text{trap}(u;.7,.9,1,1). \end{aligned} \tag{4}$$

The hedges very and more or less are the appropriate functional models of the primary terms.

In scene recognition, features were extracted from the regions of interest which included grey-level statistics, moment invariants, and texture features values from the original and segmented windows. A fuzzy pattern recognition algorithm such as fuzzy k-nearest-neighbor scheme [17] was used to produce the final class membership for each region based on the memberships of the training data and the distance (in feature space) of the sample to the training data. Given a membership m for a particular region, the linguistic confidence inherent in this value, denoted by CONF, is then modelled by trapezoidal number given by

$$\text{CONF} = \text{trap}(u;u1,u2,u3,u4), \tag{5}$$

where

$$\begin{aligned} u1 &= \max(0,m-.10), \\ u2 &= \max(0,m-.05), \\ u3 &= \min(1,m+.05), \\ u4 &= \min(1,m+.10). \end{aligned}$$

This value is then matched to the nearest (in the sense of Hamming distance) linguistic term in the database and the result is used as the

input to the system. For example, if the objective is to distinguish tanks, APCs, and false alarms, then the following subproblems are initiated to provide evidence for each pattern: target vs. false alarm, tank vs. APC, and tank vs. APC vs. false alarm. In each of these processes appropriate feature sets are chosen which can distinguish one object pattern from another.

There are several rules in the system which combine the decisions generated by the above subproblems into a final classification. These rules depend upon the results of the various subsets. For example, a typical rule for the false alarm confidence is

(RULE 20) If:

false alarm confidence is more or less high
(in target vs. false alarm)
and false alarm confidence is more or less high
(in tank vs. APC vs. false alarm)

Then:

false alarm confidence is more or less high.

In effect, we are relaxing the desire for perceived precision as found in numeric models in an effort to increase the significance and believability of the results. As an example, suppose that the false alarm confidences are high in both cases. Then the final confidence will be high. On the other hand, if the confidences are more or less high and low then the final confidence will be unknown or undecided. That is, we cannot make any decision about the false alarm confidence based on the available information because of conflicting evidence obtained by the different subproblems. By using similar rules we can also obtain the

tank and APC confidences. Note that if there is no clear winner, i.e., the resultant confidences for APC, tank, and false alarm are all unknown, then we will use a tie-breaking rule which reflects mission objectives. That is, we use the results of subproblem2 (tank vs. APC) as the final classification results for the tank and APC.

In the multisensor target recognition problem we combine evidence from several sensors to arrive at overall confidence values by fuzzy weighted averaging as discussed in [5]. If W_i and C_i denote the reliability and the confidence of the region by the ith sensor, then the overall confidence value for n sensors is given by

$$C = \frac{\sum_{i=1}^{n} W_i * C_i}{\sum_{i=1}^{n} W_i}, \qquad (6)$$

where each of the quantities is a fuzzy set.

One artifact of linguistic averaging is that the final fuzzy set has a large "tail", that is, the function rises and peaks in the original interval but trails off slower over a much larger interval. Because of this effect we perform the linguistic approximation for C by finding the Hamming distance between C and the linguistic terms in the interval [0,1] The term which provides the minimum of these calculations is chosen as the best match.

Contextual Knowledge-Based Validation

The scene has been passed through various stages of analysis when it

reaches this step in the process. Objects that have been previously stored in the reference knowledge base have been accounted for, and changes in these objects have been noticed and recorded. In some applications, this is the final step in the processing. However, if the intention is to further analyze the scene, a contextual-driven automatic object recognizer can now be used. For example, if weather conditions or time of day change, the rule base should incorporate these changes by adjusting the reliabilities of TV and FLIR sensors. A typical rule is

(RULE 5) If:

light intensity is low

Then:

the reliability if TV sensor is low

and the reliability of FLIR sensor is high.

In the area of recognition of military vehicles, scene rules are used to further interpret the information provided by scene recognition. In an image sequence, object classification confidences can be enhanced through utilization of positive evidence provided by the scene context. A typical scene rule is

(RULE 30) If:

motion detected with more or less high confidence

Then:

raise the target confidence

and lower the false alarm confidence.

In this experiment we computed the linguistic values for consistent motion in a horizontal direction as follows. Since the displacement depends on the sensor viewing angle and the distance of the object to the sensor, the center location and the prescreened window size were used to extimate the horizontal displacement between frames. Thus we computed the membership in "consistent motion" by

$$\mu_m(d) = S(d;0,wx,2wx), \tag{7}$$

where

$$S(u;a,b,c) = \begin{cases} 0, & u \le a \\ 2[(u-a)/(c-a)]^2, & a \le u \le b \\ 1-2[(u-c)/(c-a)]^2, & b \le u \le c \\ 1, & u \ge c \end{cases} \tag{8}$$

and wx and d denote the window size in horizontal direction (in the current frame) and the displacement of center points (in pixels), respectively. This membership value is then mapped to the linguistic confidence parameter as discussed in the above section.

The method to raise or to lower the linguistic confidence based on context is performed by first shifting the linguistic value to the right (toward 1) or to the left (toward 0), respectively, and then matching the result to the nearest linguistic term in the database. The amount of shifting is proportional to the power of truth value true which, in turn, depends on the degree of match between the input and the antecedent. In this experiment we computed the amount of shifting subjectively by

$$SHIFT = 0.09 * n,$$

where n denotes the power of truth value true generated by matching the

input and antecedent in rule 30. As an example, suppose that the target and false alarm confidences are both **medium** as defined in equation (4). If the confidence of the detected motion is more or less **high**, then by invoking rule 30, n will be equal 1, and as a consequence, SHIFT will be equal 0.09. Thus the target and false alarm confidences are given by

target confidence = trap(u;.49,.54,.64,.69),

false alarm confidence = trap(u;.31,.36,.46,.51).

These confidence values are then matched to the closest (in the sense of Hamming distance) linguistic term as discussed before, producing the new confidence values for the target and false alarm.

We need to point out that the number of linguistic terms in the term set is a function of the application. In a general language understanding system this problem poses an effective infinite rule and context regression [5]. However, in a limited application, this difficulty is tractable. Much depends on the expectation of the user, i.e., how expressive should the results be to satisfy the user's needs in the particular well-defined environment. The terms should not only be expressive, but the meanings should be well-understood by those employing the system. For example, an expression such as "not rather **high** to somewhat **high**" may actually increase the confusion of the situation. In this experiment the membership distributions of the linguistic values were set up subjectively to convey the meaning of natural language expressions. The values of linguistic variable **confidence** used in this experiment were (very **low**, **low**, more or less **low**, **medium**, more or less **medium**, more or less **high**, **high**, very **high**).

Implementation and Results

The fuzzy rule-based system has been implemented in the Expert

System Development Package (EXSYS) and modified by FORTRAN programs to perform fuzzy logic. The rule-based control strategy contains a forward chaining paradigm. Our prototype system uses 50 rules.

We have applied the fuzzy rule-based production system to two areas of automatic target detection and recognition. The first application involved the use of temporal sequences of FLIR images, whereas the second one concentrated the combination of evidence from both FLIR and TV images.

In the temporal case, the rule base was first trained on approximately 900 targets (tanks and APCs) and around 100 false alarms. It was then tested on a sequence of 100 frames containing two tanks and an APC during which the APC moves behind one of the tanks and into and out of a ravine. This sequence of images is considerably different from the data used to "train" the rule base. The images were prescreened, segmented, and the chosen features were extracted. The fuzzy k-nearest neighbor algorithm was used to produce the final class membership for test vectors based on the class memberships of the training data and distances of the test vector to the training data. These memberships were mapped to the linguistic confidence values as indicated. The rule base was executed and representative results are shown in Table 3. The format of the table gives the actual object under consideration, the result of various subproblems, followed by the final classification. The term unknown denotes an undecided response, that is, the rule base could not make any decision based on the available information. In tests (1), (2), and (3), the tank was classified with three different confidences and the APC was classified unknown. However, in test (4), after invoking the tie-breaking rule, the tank and APC were classified with very high

and very low confidences. In test (5), although the target confidence is very low in the target vs. false alarm subproblem, the rule base produced more or less high confidence for the both target and APC. Test (6) is an interesting case where the APC is partially occluded by the ravine. The rule base produced medium and more or less low confidences for the APC and tank, respectively. Tests (7) and (9) are two cases where the APC and false alarm are misclassified as a tank and an APC, respectively. Note that for the purpose of comparison of our method to the other schemes, target confidences were generated without incorporating the evidence obtained by the motion. However, since the APC moves, by invoking rule 30 the target confidences in tests (5) and (6) are changed from more or less high and medium to high and more or less high, respectively.

We compared our results to a rule-based system which uses Dempster-Shafer belief theory. This is the best version of the system described in Section 6.1. The comparison is displayed in Table 4. Here, the strategy has been to selectively extract groups of four features at a time and generate a set of confidences, which, in this case, would be the basic probability masses associated with the corresponding set of pattern choices. At every stage, these beliefs are combined with the ones previously obtained to give a resultant set of confidences which, after final combination, are used to make a decision about the object patterns believed to exist in the image.

In order to compare our results to the above scheme, we assigned the linguistic value more or less high as the threshold and labeled the objects as being in the class with at least more or less high confidence.

It can be seen from Table 4 that our method performed better than the belief theoretic counter part when thresholds are converted to crisp partitions. The total number of misclassifications in our method is half the number obtained by the other method. Note that even though some false alarms were misclassified as an APC, none of the targets were misclassified as false alarms. It is also important to consider the priority that exists between the tank, APC, and false alarm in a particular situation. That is, it is preferable in most situations to call an APC a tank than to call a tank an APC. As can be seen from Table 4, in our method, none of the tanks was misclassified as an APC, whereas 32 tanks were misclassified as APCs by the Dempster-Shafer method. Furthermore, the advantage of having linguistic confidence associated with the object over the crisp decisions is that a human operator or an AI routine can evaluate the input and reason about the scene. For example, since the output of inference will be unknown when there are conflicts between the pieces of evidence, this information can be used to trigger more extensive tests on the objects.

In the second situation (multisensor), the rule base was tested on two sequences of 15 frames each (FLIR and TV) and representative results are shown in Table 5. The table gives the result of classification by each sensor followed by the final classification. Since these sequences were acquired at about 2:00 p.m. and, as a consequence, the light intensity was high, the rule base assigned high and low reliabilities to TV and FLIR sensors, respectively. In each case, the results obtained by each sensor were combined using linguistic averaging to produce an overall confidence in the object. It can be seen from Table 6 that the system produced reasonable results for the objects tested. Note that

since the reliabilities of the sensors are not equal, the sensor with higher reliability (TV) will pull the final confidence towards the confidence value associated with it. Table 6 summarizes the results after labeling the objects as being in the class with at least more or less high confidence. As atmospheric conditions, time of day, and weather conditions change, the rule base can easily incorporate these changes by adjusting the reliabilities of the sensors.

Table 4. Confusion matrices for the fuzzy logic and Dempster-Shafer methods

	TANK	APC	FA
TANK	176	0	0
APC	17	49	0
FA	0	6	7

Fuzzy Logic

	TANK	APC	FA
TANK	142	34	0
APC	13	53	0
FA	0	0	13

Dempster-Shafer

FA = false alarm

Table 5. Multisensor classification results (FLIR and TV sensors)

Actual Object	TV				FLIR				Final classification			
	target	tank	APC	false alarm	target	tank	APC	false alarm	target	tank	APC	false alarm
tank (1)	MEDIUM	MEDIUM	MEDIUM	UNKNOWN	MEDIUM	MEDIUM	MEDIUM	UNKNOWN	MEDIUM	MEDIUM	MEDIUM	UNKNOWN
tank (2)	ML-HIGH	ML-HIGH	UNKNOWN	UNKNOWN	ML-HIGH	ML-HIGH	UNKNOWN	UNKNOWN	ML-HIGH	ML-HIGH	UNKNOWN	UNKNOWN
tank (3)	ML-HIGH	ML-HIGH	UNKNOWN	UNKNOWN	MEDIUM	MEDIUM	MEDIUM	UNKNOWN	ML-HIGH	ML-HIGH	MEDIUM	UNKNOWN
APC (4)	ML-HIGH	MEDIUM	ML-HIGH	UNKNOWN	MEDIUM	MEDIUM	MEDIUM	UNKNOWN	ML-HIGH	MEDIUM	ML-HIGH	UNKNOWN
APC (5)	ML-HIGH	ML-LOW	ML-HIGH	UNKNOWN	MEDIUM	MEDIUM	MEDIUM	UNKNOWN	ML-HIGH	ML-LOW	ML-HIGH	UNKNOWN

V = VERY
ML= MORE OR LESS

Table 6. Confusion matrix for multisensor case

	TANK	APC	FA
TANK	21	9	0
APC	5	10	0
FA	0	0	0

FA = false alarm

References

1. R. Crownover and J. Keller, "Fast dimension reduction that preserves undetermined data clusters," Proceedings, SPIE Conference on Advanced Signal Processing Algorithms and Architectures, San Diego, California, August 1986.

2. J. Keller, R. Crownover, J. Wootton and G. Hobson, "Target recognition using the Karhunen-Loeve transform, " Proceedings, IEEE International Conference on Systems, Man and Cybernetics, Tucson, Arizona, November 1985, pp. 310-314.

3. J. Keller, R Crownover and R. Chen, "Characteristics of natural scenes related to fractal dimension," IEEE Transactions, Pattern Analysis and Machine Intell, Vol. PAMI-9, No. 5, Sept. 1987, pp. 621-627.

4. J. Keller and D. Hunt, "Incorporating fuzzy membership functions into the perceptron algorithm," IEEE Transactions, Systems, Man and Cybernetics, Vol. PAMI-7, No. 6, November 1985, pp. 693-699.

5. J. Keller, G. Hobson, J. Wootton, A. Nafarieh and K. Luetkemeyer, "Fuzzy confidence measures in midlevel vision," IEEE Transactions, Systems, Man and Cybernetics, Vol. SMC-17, No, 4, 1987.

6. P. A. Nagin, A. R. Hanson, N. E. M. Riseman, "Region extraction and description through planning," COINS Tech Rep 77-8, Computer and Information Sciences Dept, University of Massachusetts, Amherst.

7. R. A. Brooks, R. Greiner, and T. Binford, "Progress report on a model-based vision system," Proceedings of the Image Understanding Workshop, 1978, pp.145-151 (L. S. Baumann, ed.).

8. D. P. McKeown, "MAPS: The organization of a spatial database system using imagery, terrain and map data," Proceedings: DARPA Image Understanding Workshop, June 1983, pp. 105-127.

9. D. M. McKeown, and J. McDermott, "Toward expert systems for photo interpretation," IEEE Trends and Applications, 1983, May 1983, pp. 33-39.

10. A. Rosenfeld and A. Kak, Digital Picture Processing, 2nd edition, Orlando: Academic Press, 1982.

11. R. Duda, and P. Hart, Pattern Classification and Scene Analysis, New York: Wiley & Sons, 1978.

12. K. Luetkemeyer, G. Hobson, and C. Carpenter, "Evaluation of segmentation techniques applied to prescreened areas of multi-sensor imagery," MAECON, Dayton, 1986.

13. G. Waldman, J. Wootton, G. Hobson, and K. Luetkemeyer, "A normalized clutter measure for images," Computer Vision, Graphics & Image Processing (to be published).

14. G. Hobson, and J. Wootton, "Electro optical/infrared automatic feature recognition," IRAD Tech Report, F784, Emerson Electric.

15. G. Hobson, and J. Wootton, "Electro optical/infrared automatic feature recognition," IRAD Tech Report, F785, Emerson Electric.

16. G. Hobson, and J. Wootton, "Electro optical/infrared automatic feature recognition," IRAD Tech Report, F786, Emerson Electric.

17. J. Keller, M. Gray, J. Givens, "A fuzzy k-nearest neighbor algorithm," IEEE Trans System, Man, Cybern, Vol. SMC-15, No. 4, July/August 1985, pp. 580-585.

18. J. Wootton, G. Hobson, K. Luetkemeyer and J. Keller,"The use of fuzzy set theory to build confidence measures in multisensor imagery," IEEE Applied Imagery Pattern Recognition Workshop.

19. G. Shafer, A Mathematical Theory of Evidence, Princeton: Princeton University Press, 1976.

20. A. Nafarieh, "A New Approach to Inference in Approximate Reasoning and its Application to Computer Vision," Ph.D Dissertation, University of Missouir-Columbia, 1988

7.0 Use of Context in Scene Analysis

Introduction

This section continues the discussion of uncertainty management in a rule-based system for automatic target recognition. An efficient rule-based structure should incorporate any known contextual or "a priori" information relevant to a given scene. This information pertains to situations that are exceptional or beyond the main knowledge base. Because such information is sufficiently relevant, it requires special attention to assure that it is incorporated into the rule-based decision structure. Acquired contextual information is to be used to readjust the confidences associated with the final set of choices [1-6]. In general, context in a scene refers to special information or knowledge pertaining to objects or regions in a scene or the relationship of such objects or regions. The source of this knowledge is external to the scene and this knowledge varies from situation to situation. Examples include knowledge pertaining to the number and type of objects in a scene and knowledge pertaining to whether certain objects are found together or not found together. Such specific knowledge is difficult to include in a strict rule-based structure.

The construction and application of a knowledge base in a rule-based system involves accounting for the most common occurrences or possibilities that arise in the problem domain of interest. Given a knowledge base and appropriate reasoning, a rule-based decision structure can be generated. This structure should process, evaluate and utilize as much of the available information as possible in

arriving at a "reasonable" conclusion or solution to a problem. Context information which lies outside of the existing or commonly assumed knowledge domain may not, in general, be easily incorporated into the original knowledge base. Such information, when being included in the decision-making process, could significantly influence the final conclusions.

One way to handle this problem is to modify the existing rule-based structure to accommodate or incorporate the new context information or situation. This approach could be a tedious procedure, depending on the uncertainty propagation scheme used and the rule complexity or special meta-rules used. This problem would be exaggerated if the uncertainties are propagated through a chain of rules, perhaps over a significant length of the decision tree. This would occur when, for example, the conclusion arrived at by firing a particular rule is required to satisfy part of the IF proposition or antecedent of a subsequent rule.

An alternate approach, rather than alter the original main body of the rule-based structure, is to add on peripheral or context-based rules that modify only uncertainty computations for the conclusion of rules for which the contextual information is applicable. To implement this, for every rule Rj, define a context factor,

$$Cj:[-1,1]$$

7.1 Context Factors

Based on previous discussion, the general form of a rule is given by,

PThenj = f (PIfj, PRulej);

that is, confidence in the THEN part of a rule is a function of confidence in the IF part of the rule and confidence in applying this rule.

Suppose now that it is desired to effect PThenj by a third term or parameter, Cj, the context factor associated with rule Rj. Let Cj have the following properties:

A) Cj=0 if no contextual information relevant to Rj exists

B) $C_j>0$ if relevant contextual information supports conclusion of Rj.

C) $C_j<0$ if relevant contextual information negates or disagrees with the conclusion of Rj.

Thus, the computation of the modified certainty or confidence in the conclusion of Rj, PThenj, is expressed as,

PThenj = f'(Thenj, C_j),

where f' is to satisfy the following properties:

A) P'Thenj = PThenj, if $C_j=0$ (1)

B) P'Thenj > PThenj, if $C_j>0$ (2)

C) P'Thenj < PThenj, if $C_j<0$ (3)

D) P'Thenj = 1, if C_j=0 or if PThenj = 1 (4)

E) P'Thenj = 0, if C_j=-1 or if PThenj = 0 (5)

F) For a given PThenj, P'Thenj should monotonically increase with C_j

Relevant assumptions

a) PThenj = 1: absolute indorsement in the conclusion of Rj
b) PThenj = 0: absolute refutation or rejection of the conclusion of Rj.

Thus,

PThenj, P'Thenj →[0.1] with C_j→[-1,1]

Combining Context Factors

It is likely that more than one context rule is relevant to a given rule, Rj, in the main rule base. The idea here is to first combine two or more context factors to yield a context resultant factor that is relevant to a given rule.

One approach which has been implemented, is that of sequential combination. Consider a sequence of context factors which are applicable to a given rule, Rj. First, combine the first two elements of the sequence to generate an intermediate context factor, which is then combined with the third element in the sequence to generate the next intermediate context factor, and so on until every element in the sequence has been combined so as to end up with a final (resultant) context factor for rule Rj.

Let C_{ij}: (i=1,...,m) be m context rules applicable to rule R_j.

Let C'_{kj} = intermediate context factor generated by combining the first k context rules or factors, (k=2,...,m); $C'_{1j} = C_{1j}$

Consider a function g which combines two context factors C'_{kj} and $C_{k+1,j}$ to yield a resultant factor in the form,

$$C'_{k+1,j} = g(C'_{kj}, C_{k+1,j}) \tag{6}$$

The function g is to satisfy the following desired properties.

A) $C'_{k+1,j} = C_{kj}$, If $C_{k+1,j} = 0$ (7)

B) $C'_{k+1,j} = 1$, If $C'_{k+1,j} = 1$ or
If $C_{k+1,j} = 1$ (8)

C) $C'_{k+1,j} = -1$, If $C'_{kj} = -1$ or
If $C_{k+1,j} = -1$ (9)

D) $C'_{k+1,j} = 0$, If $C'_{kj} = -C_{k+1,j}$, - cluding condition
$C'_{kj} = -C_{k+1,j} = 1$ (10)

E) $C'_{k+1,j} > \max\{C'_{kj}, C_{k+1,j}\}$, If $C'_{kj}, C_{k+1,j} > 0$ (11)

F) $C'_{k+1,j} < \min\{C'_{kj}, C_{k+1,j}\}$, I. $C'_{kj}, C_{k+1,j} < 0$ (12)

G) $C'_{k+1,j} > 0$ If $(C'_{kj} + C_{k+1,j}) > 0$ (13)

H) $C'_{k+1,j} < 0$ If $(C'_{kj} + C_{k+1,j}) < 0$.
(14)

Example Functions

The uncertainty, PThenj associated with the conclusion of a given rule Rj is modified to incorporate contextual information. First, the context is identified in terms of C_j, the context factor associated with rule Rj. Then, form a function

$f'(PThenj, C_j)$

that should satisfy properties (1) → (5).

Consider,

$$f'(PThen_j, C_j) = \begin{cases} (1+C_j)*(PThen_j)^{(1-C_j)}, & -1 \le C_j \le 0 \\ (PThen_j)^{(1-C_j)} & 0 \le C_j \le 1 \\ 1 & C_j = 1 \end{cases}$$

Now, suppose that rule Rj has M applicable context rules or factors; then the resultant context factor C_j is evaluated as follows.

Let C_{ij}: (i=1,...,m) be the M context factors applicable to rule Rj.

Let C'_{kj} be the resultar context factor that combines the first k context rules or factors. Then implement the following loop.

$C'_{1j} = C_{1j}$
DO k=1, m-1
    $C'_{k+1,j} = g(C'_{kj}, C_{k+1,j})$
ENDDO
$C_j = C'_{mj}$

The function g must satisfy properties (7)→(14). Consider the following function.

To simplify the definition of g, define:

$C'_{kj} = a$
$C_{k+1,j} = b$
$C = (a+b)/2$
$x = (1-a^2)^{1/n}$
$y = (1-b^2)^{1/n}$
$z = [x^n + y^n]^{1/2n}$

where n = same positive integer ϵ (1,2,3...).

Then, define,

$$g(a,b) = \begin{cases} a+b & \text{If } a \text{ or } b = 0 \\ |a|^{y} & \text{If } |a| \geq |b| \\ |b|^{x} & \text{If } |b| \geq |a| \\ \text{sgn}(\frac{c}{2})*[|\frac{c}{2}|^{z}], & \text{otherwise.} \end{cases}$$

7.2 The Effect of Context on Classification Results

This section briefly describes the results of a simulation that demonstrates the effect of context information on specific classification results. Consider the following example of a branch of a rule structure that can lead to the results shown in Tables 1 and 2. Here the relevant subsets of pattern classes in the "frame of discernment" are identified as follows.

S : (TANK, APC, FA, NON-TANK, NON-APC, θ)

Then, consider the following rules.

1) IF weather is 'cloudy' (Pweather = 0.90)
and if light intensity is 'low' (Pintensity = 0.85)
THEN select the FLIR sensor

2) IF sensor is "FLIR", (Psensor = ?)
THEN select feature
set F, and execute an external (P_F = ?)
algorithm to calculate confidence for the hypothesis, using
the combined evidence of 4 features.

This generates a bpa over the frame of discernment:

{m(TANK),m(APC),m(FA),m(NON-TANK),m(NON-APC),M(θ)}

with the sum of all m (•) elements = 1.

Here, bpa refers to basic probability assignment and the [m(•)]'s refer to masses corresponding to particular propositions with m(θ) being undistributed or not assigned [7-9].

Let the confidences in the validity of the given rules be as follows:

PRule1 = 0.95; PRule2 = 0.85

Using the "*" (product operator) for "∩", and the relationship,

$$P_{IF}(\text{rule } R_j) = \bigcap_{i=1}^{n} P_{IFi}(\text{rule } R_j),$$

P_{IF1} = Pweather * Pintensity = 0.765, where f is given by the product.

Also, using

$$P_{Then}(\text{rule } R_j) = f(P_{IF}(\text{rule } R_j), P_{Rule}(\text{rule } R_j)),$$

or,

$P_{Then1} = f(P_{IF1}, P_{Rule1})$, and using the * Operator for f,

$P_{Then1} = P_{FLIR} = P_{IF1} * P_{Rule} = 0.73.$

With P_{FLIR} known,

$P_{Then2} = P_F = P_{IF2} * P_{Rule2} = P_{FLIR} * (0.85) = 0.62$

Now, the validity of the evidence generated by F depends on the "proper" selection of the elements of F, which are the features extracted in the MID-LEVEL vision process as applied to a given scene.

Suppose that the above evidence is "discounted" by a factor $\alpha = h(P_F)$, where

$$\alpha = 1-(0.7+0.3*P_F) = 1-0.89 = 0.11, \quad [7]$$

Now, consider two applicable context rules for rule R_2

CR1: IF number of objects detected is less that number of objects expected (in scene)

THEN increase confidence in any relevant object subsequently located in the scene ($C_{21} = 0.5$).

ELSE decrease confidence in any additional objects identified ($C_{21} = -0.5$)

CR2: IF object detected = object expected at a given location,

THEN increase confidence in an object being detected at that location ($C_{22} = 0.8$)

Suppose:

1) number of objects detected is 4;
2) number of objects expected is 3;
3) object expected in a given window (location) is a TANK;
4) the corresponding F yields a bpa vector $\underline{M}$,

$$\underline{M} = [0.7\ 0.1\ 0.0\ 0.0\ 0.0\ 0.2]^T .$$

This bpa vector is then "discounted" by α to form,

$\underline{M}' = [0.62\ 0.9\ 0.0\ 0.0\ 0.0\ 0.29]$. (Discounting is discussed below.)

The maximum mass of $\underline{M}$ is 0.62,which is associated with the class TANK, and thus the intermediate classification by the rule base is TANK with a confidence of 0.62.

Context rule C_1 yields $C_{21} = -0.5$, while context rule C_{R2} yields $C_{22} = 0.8$. Therefore, the resultant context factor for rule R_2 using the function g defined above is given by,

$$C_2 = g\ (-0.5,\ 0.8) = 0.33$$

Then, the final confidence in the TANK classification using function f described above is given by,

$$conf(TANK) = f(0.62,\ 0.33) = 0.73.$$

Table 1. Classification results - Each entry indicates the proportion of correct classifications from a sampled set of available data.

	bpa's using π-function				bpa's using Fuzzy Integral			
	TANK1	TANK2	APC	FA	TANK1	TANK2	APC	FA
SIM1: polling	8/8	14/17	2/24	13/13	17/17	17/33	47/66	13/13
SIM9: global D-S	8/8	11/17	8/24	13/13	17/17	17/33	53/66	13/13

Table 2. Simulation showing effect of "context".

No	SENSOR	C.M. used for TV	C.M. used for TV	# Of objdet < # Of objexp	CLSDET =CLSEXP	C	C'
1	TV	NO	NO	N.A.	N.A.	0.26	0.26
2	TV	YES	NO	N.A.	N.A.	0.26	0.05
3	TV	NO	YES	N.A.	N.A.	0.26	0.43
4	FLIR	NO	NO	YES	N.A.	0.21	0.73
5	FLIR	NO	NO	N.A.	YES	0.21	0.63
6	FLIR	NO	NO	NO	N.A.	0.26	0.07
7	FLIR	NO	NO	NO	YES	0.26	0.32
8	TV	YES	YES	N.A.	N.A.	0.26	0.28
9	FLIR	NO	YES	NO	NO	0.26	≈0.0
0	FLIR	YES	NO	YES	YES	0.26	0.87

where,

C.M. : Counter Measures
C : Initial Confidence in classification
C' : Confidence (C) modified with context
OBJDET : (# of) Objects detected so far
OBJEXP : (# of) Objects expected in image
CLSEXP : Object expected at current location
CLSDET : Object detected at current location
N.A. : Not Applicable

Discounting and Disparity

Consider two sources, S_1 and S_2 of information, which can generate a rule base for a rule-based decision structure to provide solution in a given problem domain. Suppose that S_1 is more noise or error prone than S_2. For a given problem, using source S_1 yields a answer A with certainty.8 while using source S_2 yields answer B with certainty 0.6. Because source S_1 is more noisy or error prone than source S_2, it is less reliable. In general, if the answers provided by sources S_1 and S_2 are to be used together, one must compensate the final conclusion of the decision structure based on source S_2 for the effect of information from source S_1.

Discounting

Given a belief function Bel, defined over a-Frame of Discernment S: $\{A_1, A_2, \ldots, A_n\}$, discount or degrade it by discounting the belief in every proper subset $A \subset S$ by a discount rate α. Thus, reduce m(A) to $(1-\alpha)*m(A)$. Because the sum of all bpa's defined over S add to 1, the total sum of all the discounted evidence must be added to $m(\theta)$, the undistributed or non-specific belief. The influence of this source of evidence on the final outcome is now discounted or reduced (because $m(\theta)$ increases) [7,10].

Example

Consider a pattern analysis problem, where S: (A,B) represents two pattern classes. For two equally reliable sources S_1 and S_2 that provide independent judgements, assume

S_1: $m_1(A) = 1$; $m_1(B) = 0$; $m_1(\theta) = 0$

S_2: $m_2(A) = 0$; $m_2(B) = 1$; $m_2(\theta) = 0$.

For two independent sources of evidence S_1 and S_2, define belief functions $(Bel)_1$ and $(Bel)_2$, respectively for a given proposition X. Then, using Dempster's rule to combine evidence from sources S_1 and S_2 to generate a resultant belief Bel(X), as given by,

$$k = 1 - \sum_{B \cap C = \phi} m_1(B) * m_2(C)$$

$$m(A) = \sum_{B \cap C \subset A} m_1(B) * m_2(C) / k$$

$$Bel(X) = \sum_{A \subset X} m(A),$$

where k represents the measure of conflict between (Bel) and $(Bel)_2$, as it accounts for the belief jointly allocated to ϕ by the two sources.

Here, because $m_1(\theta) = m_2(\theta) = 0$, one arrives at an undefined situation. To account for an apparent unreliability of sources S_1 and S_2, discount them by α so that

$$S_1: m_1'(A) = (1-\alpha);\ m_1'(B) = 0,\ m_1'(\theta) = \alpha$$

$$S_2: m_2'(A) = 0;\ m_2'(B) = (1-\alpha);\ m_2'(\theta) = \alpha$$

The resultant bpa's (or mf's) after combination are,

$$S_1|S_2: mf(A) = (\alpha-\alpha^2)/(2\alpha-\alpha^2) = mf(B)$$
$$mf(\theta) = \alpha^2/(2\alpha-\alpha^2)$$

One can show that (in the limit $\alpha \to 0$) $mf(A) = mf(B) = 0.5$ and $mf(\theta) = 0$.

Suppose that source S_1 is unreliable or perhaps only more unreliable than source S_2; then discounting just source S_1,

$$S_1: m_1'(A) = (1-\alpha);\ m_1'(B) = 0;\ m_1'(\theta) = \alpha$$

$$S_2: m_2'(A) = (1-\alpha);\ m_2'(B) = 1;\ m_2'(\theta) = 0.$$

after combining,

$$S_1|S_2\text{: } mf(A) = 0,\ mf(B) = 1;\ mf(\theta) = 0$$

If one cannot determine which source is more unreliable, first combine information from sources S_1 and S_2, then discount the results. This will yield,

$$S_1|S_2\text{: } mf(A) = (1-\alpha)/2 = mf(B),\ mf(\theta) = \alpha.$$

Disparity

Suppose that a source of evidence or information, S, derives its information from a number of "subsources" : $\{S_1,\ldots,S_m\}$, such as α consisting of different sets of features used by a given classifier algorithm. Each such set would contribute a set of beliefs regarding an observed pattern. In general, one would assume that these beliefs do not completely agree. One way to handle this case is to output a vector, each component being associated with a degree of uncertainty.

Consider a set of belief functions $\{(Bel)_1,\ldots, (Bel)_m\}$ associated with m subsources of a given source S. Each $(Bel)_i$ vector has n components, i=1,..., n with n = total number of proper subsets of S.

Then, $(Bel)_i = [a_{i1},\ldots, a_{in}]^T$,

where $a_{ij} = (Bel_i(A_j))$, $A_j \subset S$

Consider a vector R,

$\underline{R} = [r_1, \ldots, r_n]^T$, where

$r_k = \sigma_k$, the standard deviation of $\{a_{1k}, \ldots, a_{mk}\}$, the set of all the k^{th} elements of each belief vector.

Now, the effective length of vector R in the p-norm sense is given by,

$$L_R = \left[\sum_{k=1}^{n} |r_k|^p \right]^{1/p} ;$$

For each r_k, it can be shown that

$$r_k|_{max} = \begin{cases} 0.5 * (\frac{m}{m-1})^{1/2}, & m \quad n \\ 0.5 * [\ \frac{m+1}{m}\ ]^{1/2}, & m \text{ odd} \end{cases}$$

so that,

$$L_r|_{max} = \begin{cases} n^{1/p} * (m/4(m-1))^{1/2} \\ \\ n^{1/p} * (mfl)/4m)^{1/2} \end{cases}$$

(n is dimension of R).

Now, define the disparity of the m sub-sources $\{S_1, \ldots, S_n\}$ as,

$$d(S_1, \ldots, S_m) = L_R/L_R|_{max} \rightarrow [0,1].$$

Using this measure to generate α between, say, [0,0.3],

$$\alpha = 0.3*d[S_1, \ldots, S_m]$$

The beliefs resulting from combining the evidence of the m subsources of S would be discounted by α, as defined above. This discussion of disparity represents only a preliminary effort. Further study is necessary to establish its efficacy. A disparity measure could be applied to other methods of uncertainty propagation. For example, it could be used to generate weights associated with each of the main sources of evidence before they are combined using the method applicable to a given scheme.

The Direct Combination of Context: A Preliminary Investigation

Suppose that the context factors were allowed to assume values over $(-\infty,\infty)$ instead of over [-1,1] as was proposed earlier. For a given rule R_j, define:

C_{kj}: k^{th} context factor (describing or resulting from the k^{th} context rule) applicable to rule R_j.

N_j: total number of context rules applicable to rule R_j.

C'_j: algebraic sum of the N_j context factors applicable to rule R_j.

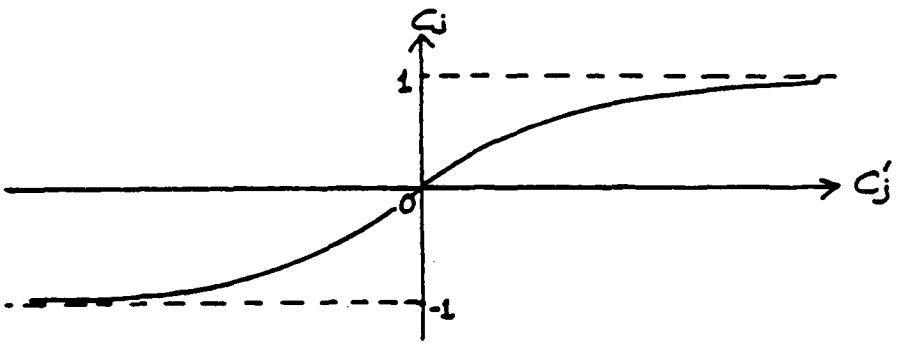
Then,

$$C'_j = \sum_{k=1}^{N_j} C_{kj} \rightarrow (-\infty,\infty)$$

Now, in order to incorporate this approach into rule R_j, it has been necessary to restrict the resultant context factor, C_j, to be defined over [-1,1]. In order to retain this property, one can define a one-to-one mapping between $C'_j:(-\infty,\infty)$ and C_j: (-1,1). By doing this, one will end-up with a $C_j:[-1,1]$, which can then be used in

$$PThenj = f(PThenj, C_j).$$

Such a mapping can be typically represented by the following sketch:



in particular, let $C_j' = \tanh (C_j/C_o)$, where C_o is a "tuning" parameter which adjusts the slope or "sharpness" of the mapping near the origin.

For this mapping, note that

a) for $C_j' \geq 5\ C_o$, $C_j = 1.0$

b) for $C_j \leq -5C_o$, $C_j = -\ 1.0$

Thus, the assignment of relative values to the individual C_{kj}'s need not be as "arbitrary" as it otherwise might be if normalization were retained.

For a fixed value for C_o, the individual C_{kj} values can be added together directly before the sum is mapped, one-to-one onto C_j'. One can also estimate the effect of an individual C_{kj} value on the

resultant value of C_j' and then on C_j. This method of combining context factors preserves the properties of commutativity and associativity.

REFERENCES

1. Bhatnagar, R. K. and Kanal, L. N., "Handling uncertain information: A review of numeric and non-numeric methods", Uncertainty in Artificial Intelligence, Kanal, L. N. and Lemmer, J. F., Eds. (New York: North Holland, 1986).

2. Bonissone, P. P. and Tong, R. M., "Editorial: Reasoning with uncertainty in expert systems", Int. J. Man-Machine Studies, (1985) 22, pp. 242-247.

3. Cohen, P. R., "Heuristic reasoning about uncertainty: An artificial intelligence approach", Uncertainty in Artificial Intelligence, Kanal, L. M. and Lemmer, J. F., Eds. (New York: North Holland, 1986), pp. 37-42.

4. Lesmo, L., Siatta, L. and Torassu, P., "Evidence combination in expert systems", Int. J. Man-Machine Studies (1985) 22, pp. 311-313.

5. Mamdani, E. H., and Efstathiou, H. J. , "Higher-order logics for handling uncertainty in expert systems", Int. J. Man-Machine Studies (1985) 22, pp. 290-292.

6. Martin-Clouaire, R. and Parde, H., "On the problems of representation and propagation of uncertainty in expert systems", Int. J. Man-Machine Studies (1985) 22, pp. 254-257 and 260-262.

7. Shafer, G., A Mathematical Theory of Evidence, (Princeton University Press, Princeton, N. J. 1976).

8. Shafer, G., "Belief functions and possibility measures", in The Analysis of Fuzzy Information, 2 (CRC Press, 1987).

9. Shafer, G. and Logan, R., "Implementing Dempster's Rule for Hierarchical Evidence", Artificial Intelligence (1987) 33, pp. 273-278.

10. Smets, P., "Information content of an evidence", Int. J. Man-Machine Studies (1983), pp. 35-38.

APPENDIX

The appendix contains related documents produced as a result of this grant.

Contents:

1. J. Keller, S. Chen, and R. Crownover, "Texture description and segmentation using fractal geometry", Comput. Vision, Graphics, Image Proc., accepted for publication.

2. S. Chen, J. Keller, and R. Crownover "On the calculation of fractal features from images:, IEEE Int. Conf. on Computer Vision, Dec., 1988, under review.

3. S. Chen, J. Keller, and R. Crownover, "Shape from fractal geometry", Artificial Intelligence, under review.

4. H. Tahani and J. Keller, "Information fusion in computer vision using the fuzzy integral", IEEE Trans. Syst., Man, Cybern., under review.

5. R. Corwnover, "A least squares approach to linear discriminant analysis", SIAM J. Scientific, Stat. Computing, under review.

6. A. Nafarieh, "An approach to inference in approximate reasoning and its applications to computer vision", Ph.S. dissertation, University of Missouri-Columbia, August 1988 (Chapters 4, 5, 6, and appropriate appendices. Chapters 1-3 contain background material easily obtainable in standard references.)